RICE UNIVERSITY

C_{60} Amino Acids and Peptides

by

T. Amanda Strom

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Andrew R. Barron, Chair
Charles W. Duncan, Jr. – Welch Chair of
Chemistry and Professor of Materials Science
Associate Dean of Industrial Interactions and
Technology Transfer

Zachary Ball
Assistant Professor of Chemistry

Rebecca Richards-Kortum
Stanley C. Moore Professor of Bioengineering

HOUSTON, TEXAS
AUGUST 2010

ABSTRACT

C_{60} Amino Acids and Peptides

Since the discovery of the buckyball in 1985, researchers have imagined its potential in fields ranging from materials science to medicinal chemistry. The unique size, shape and hydrophobicity of C_{60} fullerene endow it with the ability to interact with biological superstructures such as enzymes and membranes making it attractive as a potential pharmacophore. In this regard, we have developed a new, simple route to water soluble fullerene amino acids, both alkyl and aryl, through the dipolar addition of azido starting materials. The synthesis of our phenylalanine derivative, including the chromatographic purification, requires only one day for its completion. We have subsequently used our C_{60} phenylalanine derivative in the synthesis of a series of C_{60} peptides for the purposes of enzyme inhibition, specifically human immunodeficiency virus Type 1 protease, a critical viral enzyme responsible for the maturation of the virus and a popular target of medicinal chemists. We have demonstrated the ability of our C_{60} amino acids and peptides to inhibit HIV-1 PR in a cell-free fluorescence based assay at low nanomolar concentrations.

Graphite, or specifically graphene, has recently come to the forefront of nanomaterials research due to it similar scale, properties, and reaction pathways as other more costly carbon nanostructures such as carbon nanotubes. We have demonstrated the high yield functionalization of graphitic starting materials through the thermal decomposition of azido amino acids to their corresponding nitrene. The result is an inexpensive, highly functionalized, carbon based scaffold.

Acknowledgments

In the course of my work with fullerene amino acids and peptides, many collaborations were formed, and so I will take this opportunity to thank them. I would like to thank Dr. Nancy Monteiro-Riviere and her group at North Carolina State University for their toxicological and membrane penetration work that has advanced our knowledge of fullerene amino acids and peptides. Dr. Jianhua Yang and co-workers at Texas Childrens Cancer Center also performed toxicological experiments for us. In addition to Dr. Jianhua Yang, his brother Dr. Jianzhong Yang was the original developer of the first fullerene amino acid to come out of our laboratory. I thank him for his time, in person and in emails, over the years.

During the graphene project, I had many collaborators for the characterization of my nanomaterials. I would like to thank Dr. Noe Alvarez for the TEM micrographs and Eoghan Dillon for the STEM, AFM, and XPS characterization.

I would like to thank my collaborators Dr. Serdar Durdagi at the University of Calgary and Dr. Manthos Papodopolous at the National Hellinic Reasearch Foundation in Greece for their computer modeling program that finally gave me the opportunity to visualize my molecules in their intended application. In addition to the computational program, we had many fruitful and motivating conversations about our work and eventually became good friends also, even if it was only through email. I would also like to thank Dr. Claudiu Supuran at the University of Florence in Italy and Dr. Thomas Mavromoustakos at the University of Athens in Greece for all of the early biological assays and medicinal chemistry discussions during the early stages of the HIV inhibitor project. Taken together, this was the most productive and exciting collaboration I have ever been a part of and it was quite inspirational to have all of our work crossing an ocean; it truly was an international, multidisciplinary team. I have only one regret, I never did get the chance to personally deliver my samples as I had hoped.

I would like to thank Dr. Zachary Ball, who sits on my committee, the first professor to truly challenge me, more so than ever before, with his advanced organic chemistry course. He devoted untold hours to his students that first year, regarding coursework, research, and sometimes college football. We had many conversations that were appreciated very much. I also have to thank the past and present members of the Shared Equipment Authority that have been vital to my research. I managed to pick up quite a bit from Dr. Larry Alemany and especially Dr. Sandy Yates in the NMR and mass spec labs respectively. And while I never used his instrument, Richard Crouse and I spent many hours together trying to keep the mass spectrometry lab open. I truly appreciated his selflessness considering how much of an extra workload that was for him.

Most importantly I would like to thank my group, past and present, but particularly the five I came in with: Chris, Johnathan, Nadjmeh, Huma, and Chris. And I cannot forget about the Irish contingent that has been a constant fixture in our lab: Al, Colm, and Caoimhe. We all have had countless discussions, arguments, tears, poker parties and bar-b-ques. My lab mates were the first people I ever met that were as excited by science as I was and have been there for the most important moments of my life. I miss the ones that have gone on to bigger and better now, but I will never forget them. They are great friends and were great co-workers.

Not to be left out, I would like to thank my boss, Dr. Andrew Barron. He always let me work as independently as I wanted and was always there in times when I needed him. No matter what, I always knew that Andy was on my team and would support me in any arena. That kind of allegiance is rare in an advisor I think. Not to mention, his involvement with an Irish exchange program introduced me to my husband (and he let him stay).

My biggest thanks go to Eoghan Dillon, my husband, who sits next to me in the office and has the bench across from me in the lab. How we have not killed each other is

a mystery to me. Quite the opposite, he has gotten me through those very stressful times that are sometimes a part of research. I honestly can't believe he didn't try to escape during this thesis writing portion of my academic career. He is my best friend and the only person who can talk reason into me. I don't know how I would have gotten through my candidacy without him. Finally, I would like to thank my parents, who still wonder if I'll ever get out of school. They have supported me in more ways than one over the years. So, thank you to all the people that have been a vital part of my time at Rice.

Table of Contents

List of Figures

Introduction

Chapter 1

List of Tables

Chapter 2

Chapter 3

Abbreviations

A	ampere
Å	angstrom
AFM	atomic force microscopy
atm	atmosphere (101,325 Pa)
Baa	Bucky amino acid, fullerene phenylalanine
Boc	*tert*-butyloxidecarbonyl, $-C(O)OC(CH_3)_3$
C_{60}	Buckminsterfullerene, [60] fullerene
°C	degrees centigrade
ca.	circa, approximately
cm	centimeter, 10^{-2} m
°	degrees
δ	delta, chemical shift (NMR)
DCM	dichloromethane
DIEA	diisopropylethylamine
DMF	N, N-dimethylformamide
DNA	deoxyribonucleic acid
ESI-MS	electrospray ionization-mass spectroscopy
Eq.	equation
EtOH	ethanol, CH_3CH_2OH
Et	ethyl, $-CH_2CH_3$
et al.	*et alia*, and others
FITC	fluoreseinisothiocyanate
Fmoc	9-fluorenylmethoxyloxide
g	gram
h	hour

HBTU	1H-Benzotriazolium, 1-[bis(dimethylamino)methylene]-, hexafluorophosphate(1-), 3-oxide (9CI)
HOBt	N-hydroxybenzotriazole
HPLC	high-performance liquid chromatography
Hz	hertz
i.e.	*id est*, that is (to say)
IR	infrared spectroscopy
J	Joule
K	kelvin
λ	lambda, wavelength
L	liter(s), ligand
m	meta
M	molar
μm	micrometer(s), 10^{-6} m
μL	microliter(s), 10^{-6} L
m	multiplet(NMR), medium(IR), meter(s)
Me	methyl, $-CH_3$
mg	milligram(s), 10^{-3} g
mL	milliliter(s), 10^{-3} L
mM	millimolar, 10^{-3} M
MS	mass spectroscopy
MW	molecular weight
min	minute(s)
mol	mole(s)
MeOH	methanol, CH_3OH
mmol	millimole(s), 10^{-3} mol
MALDI-MS	matrix-assisted laser desorption ionization mass spectroscopy

nm	nanometer, 10^{-9} m
NMR	nuclear magnetic resonance spectroscopy
o	ortho
ppm	parts per million
PyBOP	1-H-benzotriazole-1-yloxytripyrrolidinophosphonium hexafluorophosphate
rpm	revolutions per minute
RNA	ribonucleic acid
s	singlet (NMR), strong (IR), second (s)
SEM	scanning electron microscopy
SPPS	solid phase peptide synthesis
STEM	scanning tunneling electron microscopy
SWNT	single walled carbon nanotube
TEM	transmission electron microscopy
TFA	trifluoroacetic acid
TIPS	triisopropylsilane
TOF	time of flight
UV-vis	ultraviolet/visible spectroscopy
VSI	vertical scanning interferometry
XPS	X-ray photoelectron spectroscopy

Introduction

Amino acids are one of the most basic and important building blocks of life. Most were discovered over a century ago when proteins were boiled in hot acids and the more simple products of the reaction crystallized out of solution.[1] The first amino acid discovered in nature and then isolated by a group of French chemists was from asparagus roots and was so called asparagine.[2] Not to be left out, Japanese chemists soon led the way in determining the identity and properties of amino acids after the discovery that glutamate was responsible for the savory umami taste.[3] During the time of their discovery and characterization, amino acids and the proteins they came from were considered quite baffling substances to early chemists. Figure I.1 shows the generic structure of an amino acid composed of a tetrahedral, α-carbon center attached to an amino group, a carboxyl group, a hydrogen atom, and an R side chain, which imparts the overall structure with its unique properties. At biological pH (7.4) an individual amino acid exists as a zwitterion, charged at both ends but overall neutral.

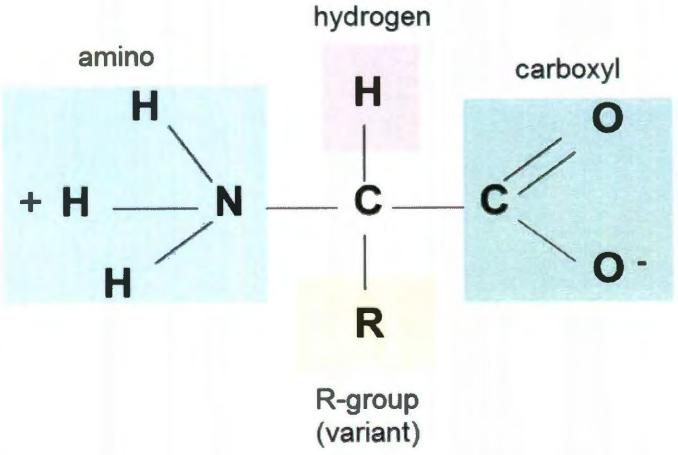


Figure I.1. The generic base structure of all of the twenty naturally occurring amino acids.[4]

There are twenty, naturally occurring common amino acids found in proteins with distinct properties due to their structural difference in their R sidechain. Of the twenty naturally occurring amino acids, ten are essential for human life.[5] Figure I.2 shows the structures all of the amino acids along with their classification as either: nonpolar or aliphatic, polar but neutral, acidic or negatively charged, and basic or positively charged. The naturally occurring amino acids found in proteins are α-amino acids, meaning the carboxy, amino, R sidechain, and hydrogen moieties are attached to the alpha carbon imparting chirality to the structure (except glycine where R is hydrogen). As may be evident by their structures, the identity of the R sidechain controls the properties of the amino acid. For instance, the non-polar or aliphatic amino acids are hydrophobic in nature and therefore tend to reside in the hydrophobic core of a protein or within lipid membranes.[5]

The discovery that amino acids are asymmetric or chiral necessitated the development of a suitable nomenclature system. As a result both the D/L and R/S systems were developed. The D/L system is based on the principal that enantiomers, non-superimposable mirror images, rotate plane polarized light in different directions but to the same degree.[5] Clockwise rotation (+) of the incident light is designated as dextrorotatory whereas counterclockwise rotation (-) is referred to as levorotatory; however, this is not what the D/L system refers to. The D/L system uses D and L glyceraldehyde as a reference for all other carbon based molecules.[5] Given that all protein-derived amino acids are of the L configuration but some are dextrorotatory and some are levorotatory, this naming system can be confusing. In the R/S system, priorities are given to the chiral carbon substituents based on atomic number. Atoms with higher atomic numbers are given the higher priority.[5] The R/S system is less ambiguous and therefore will be used in the remainder of this discussion.

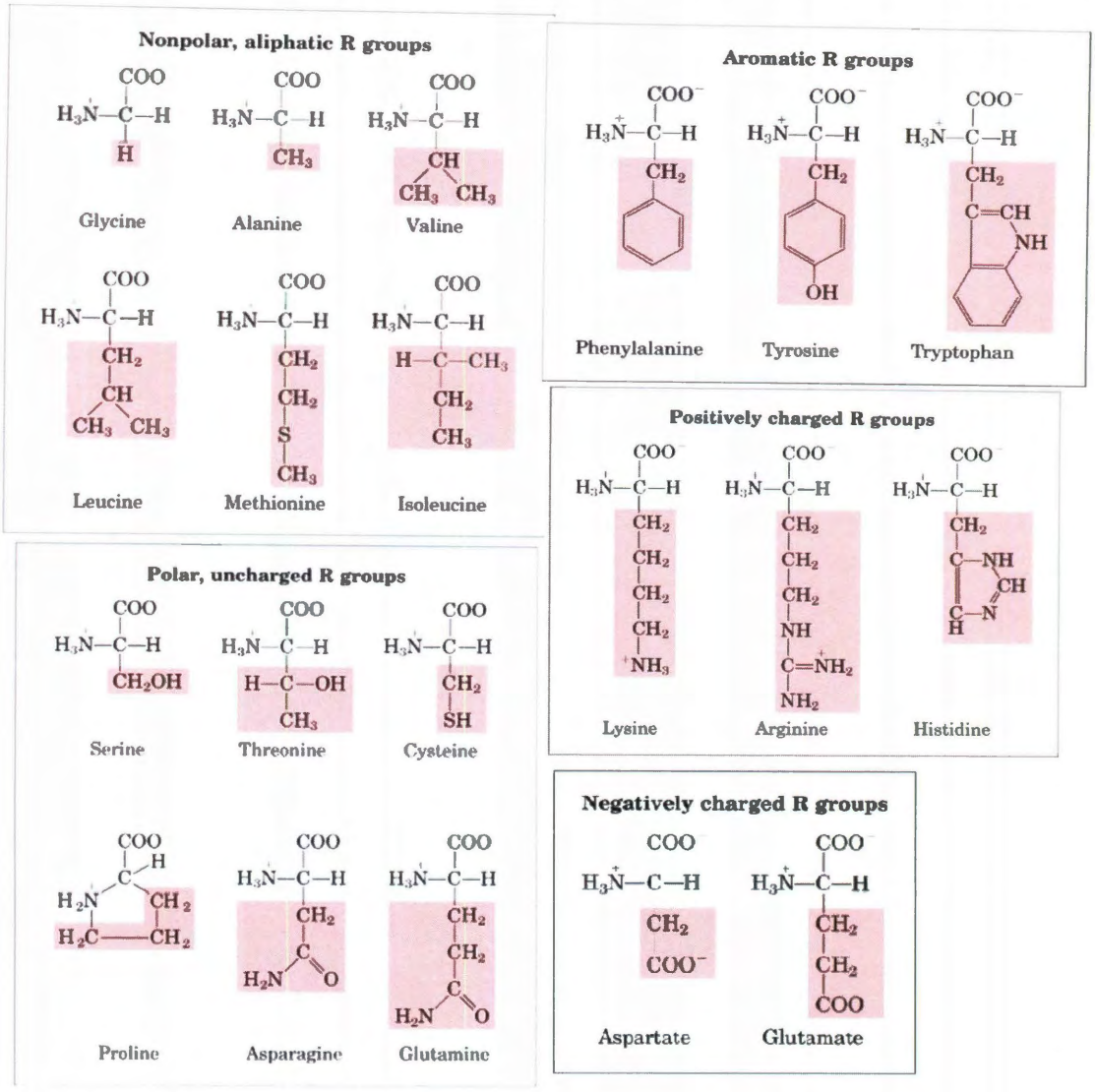


Figure I.2. The structures of the twenty common amino acids along with their classification. The R-side is highlighted in pink.[6]

In addition to the twenty common amino acids, there are several that occur only rarely in proteins, such as hydroxylysine and hydroxyproline, which are found in collagen and gelatin proteins.[5] Other minor alterations to the common amino acid structures include methylation in certain muscle proteins and phosphorylation of amino acids

involved in cell growth and regulation.[5] All of these variations on the common amino acid structure are found in proteins; however, there are still many biochemically important amino acids and derivatives not found in proteins. Some examples are the neurotransmitters γ-aminobutyric acid (GABA) and serotonin, and the hormone epinephrine.[5]

Amino acids polymerize in a head to tail fashion through their identifying functional groups, NH_2 and COOH, in a condensation reaction forming an amide bond, also referred to as a peptide bond (Figure I.3). The nitrogen is in resonance with the carboxy group preventing free rotation around the peptide bond and resulting in the partial double bond character of peptide bonds. The structures formed by the polymerization of amino acids are designated as peptides when they are short polymers, polypeptides when several dozen amino acids are present, and proteins when they consist of one or more polypeptide chains.[5]

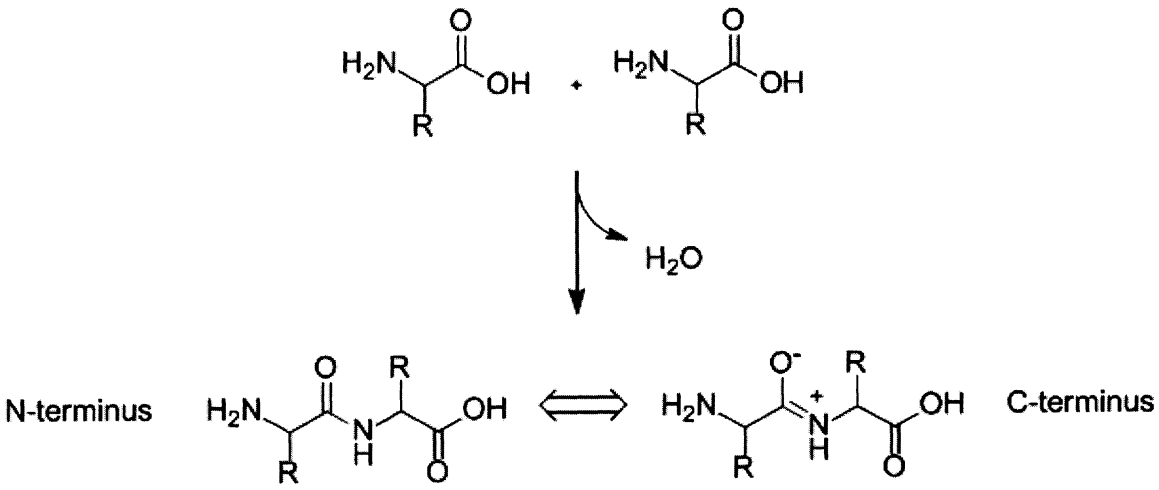


Figure I.3. The polymerization of amino acids into peptides is presented above with the peptide bond highlighted in red.

A protein has several levels of structural organization. When composed of one peptide chain they are monomeric, when composed of multiple peptide chains they are

referred to as multimeric. When a protein is composed of two identical monomers, it is called a homodimer. The architecture of a protein can be quite complex, but can be clarified by designating various levels of organization. The primary structure is simply the sequence of amino acids present. Non-covalent interactions, such as H-bonding, between polypeptide chains form the secondary structure. The overall 2° structure is organized into new specific structures such as α–helices and β–sheets having the specific properties imparted by the amino acid sequence. When the polypeptide chains bend and fold they form a compact, three-dimensional structure designated as the tertiary structure. When multiple 3° structures interact with each other they form the quaternary protein structure. The conformation of a protein refers to its overall protein architecture and it quaternary structure is the functional unit of almost all cellular functions. [5]

One subset of the functional quaternary protein structures, and by far the largest, is enzymes. Enzymes function to catalyze biological reactions and do so with enviable efficiency from a chemist's point of view. Enzymes can improve biological reaction rates by as much as 10^{16} times the uncatalyzed reaction.[5] For example, DNA synthesis requires the decarboxylation of orotidine monophosphate (OMP). The uncatalyzed half-life of the decarboxylation of OMP is approximately 78 million years, while the enzyme catalyzed reaction half-life is 18 milliseconds.[7] Enzymes are usually conveniently named according to the nature of the reaction they catalyze. For instance the enzyme responsible for catalyzing the decarboxylation of OMP is called OMP decarboxylase. Enzymes are not just essential to sustaining life in the human body but also in sustaining disease.

When a person is infected with a virus such as human immunodeficiency virus (HIV) or hepatitis C virus (HCV), otherwise healthy cells are reprogrammed by the virus to produce viral proteins/enzymes. The viral enzymes function for the virus in the same manner as they do in healthy cells by catalyzing otherwise kinetically unfavorable reactions. A virus is not always the reason for dysfunctional enzymes; sometimes the body overproduces an enzyme leading to disease or enzyme point mutations cause it to

malfunction.[7] For these reasons, enzymes have become a common target for medicinal chemists. In fact, in 2002, almost half of all marketed small molecule drugs functioned as enzyme inhibitors.[7]

It was widely believed that the sequencing of the human genome would bring about bountiful targets for drug therapy. In fact, of the some 30,000 proteins encoded for by the human genome, only about 10% are considered to be "druggable".[7] The term "druggable" refers to the ability of a target to fold into conformations able to interact specifically with small drug-like molecules.[8] Of those 3000 "druggable" targets, only about 600 - 1500 of them are considered to be "disease modifying", with a large percentage of them being enzymes.[7] Given that in many cases the substrate of a given enzyme is known, medicinal chemists are essentially provided with a roadmap to inhibition. Hence the popularity of enzymes as targets for drug therapy. However, many proteins interact only with other proteins, not small drug-like molecules. These targets were widely considered by pharmacologist to be unlikely targets for inhibition by small molecule drug therapy.[8] In those cases it became necessary to tap a source of otherwise non-classical drug molecules, such as peptides and antibodies, as pharmacophores.

Drugs built on a peptide base are not without their drawbacks. Peptides tend not to be able to survive the acidic environment of the gastrointestinal tract and therefore must be delivered in some other, usually uncomfortable, means such as injection.[9] Regardless, there are countless peptide drugs on the market today. A few examples include insulin, the first and most commonly prescribed peptide drug, for the treatment of diabetes, interferon for the treatment of multiple sclerosis and hepatitis, and erythropoietin for the treatment of the chronic anemia experienced by patients in renal failure.[9] While peptides as drugs do present some limitations, they also possess many advantages.

If protein interfaces determine the interactions between a protein and its substrate, then logically a compound mimicking that interface should compete with the natural

substrate. Peptides as enzyme inhibitors have the benefit of a structure that is inherently compatible with a protein's active site. As it turns out, a relatively small number of amino acids typically govern the interactions of protein-protein interfaces; among them are proline (Pro), isoleucine (Ile), tyrosine (Tyr), tryptophan (Trp), asparagine (Asn) and arginine (Arg).[8] What this means is that only a small number of amino acids may be critical for molecular recognition of a peptide inhibitor. If problems with stability, delivery, and specificity cannot be overcome, a peptide inhibitor can also serve as a guide for the synthesis of non-peptidic inhibitors.

One issue of many drug molecules is their inability to cross a membrane, whether that is a cell membrane or the blood-brain barrier. In fact a set of guidelines called Lipinski's rule of five was established and widely recognized for evaluating the "drug likeness" of compounds.[10] The main goal of these guidelines is to predetermine if a drug candidate will be orally bioavailable and whether it will have the ability to permeate cell membranes. But there is another solution to this problem of permeability that lies in the field of nanotechnology.

Since its discovery in 1985[11] and bulk production in 1990,[12] C_{60} fullerene has captured the imagination of researchers. The given name of the unique molecule is Buckminsterfullerene and the IUPAC method of naming distinguishes it as hentriacontacyclo[29.29.0.$0^{2,14}$.$0^{3,12}$.$0^{4,59}$.$0^{5,10}$.$0^{6,58}$.$0^{7,55}$.$0^{8,53}$.$0^{9,21}$.$0^{11,20}$.$0^{13,18}$.$0^{15,30}$.$0^{16,28}$.$0^{17,25}$.$0^{19,24}$.$0^{22,52}$.$0^{23,50}$.$0^{26,49}$.$0^{27,47}$.$0^{29,45}$.$0^{32,44}$.$0^{33,60}$.$0^{34,57}$.$0^{35,43}$.$0^{36,56}$.$0^{37,41}$.$0^{38,54}$.$0^{39,51}$.$0^{40,48}$.$0^{42,46}$]hexaconta-1,3,5(10),6,8,11,13,(18),14,16,19,21,23,25,27,29(45),30,32,(44),33, 5(43), 36,38(54),39(51),40(48),41,46,49,52,55,57,59-triacontaene.[13] But for the purposes of this discussion the more convenient names Buckyball, fullerene, and C_{60} refer exclusively to C_{60} fullerene and are used interchangeably.

Consisting entirely of fused hexagons and pentagons, the class of molecules called fullerenes have a curved shaped, as a result of the pentagons, and are the only allotropes of carbon that are soluble in organic solvents at room temperature.

Specifically, C_{60} has twelve pentagons and twenty hexagons giving it a perfectly spherical shape with all of the pentagons isolated by hexagons (Figure I.4). The bonds of [6,6] junctions, where the electron density is centered, are shorter than the bonds of [5,6] junctions.[13] The strained shape of C_{60} along with its electron deficiency and conjugated system of double bonds make it an extremely reactive molecule in nucleophillic additions, cycloadditions, radical additions, reductions and many others.[13] The outer diameter of C_{60} with its associated electron cloud is 10.34 Å or 1.034 nm. Some might call C_{60} the original nanoparticle.

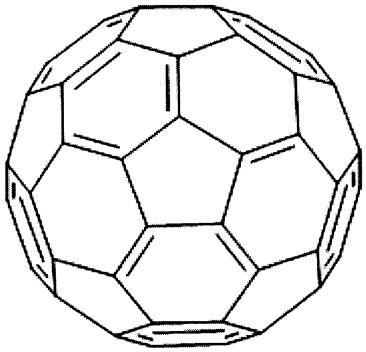


Figure I.4. The structure of C_{60} fullerene (Buckyball).

The unique 1 nm size, spherical shape and electrostatic properties of C_{60} have sparked interest in the novel molecule spanning multiple scientific disciplines from materials science to medicinal chemistry. The many known and potential applications of C_{60} are too broad and diverse a topic for one discussion; therefore, this one will focus on some of the biological applications of C_{60} derivatives.

One hurdle early fullerene chemists had to overcome in the study of the biological applications of fullerenes was the aqueous solubility of C_{60} or rather the lack there of. However, progress in the field has all but eliminated the problem of solubility through derivitization with polar substituents, dendrimers, and biomolecules, opening the door to

a wide range of biological applications including photodynamic therapy, DNA cleavage, and enzyme inhibition.[14]

The size and shape of C_{60} endow it with the unique ability to interact with biological superstructures such as enzymes and receptors.[15] Friedman was the first to propose the fit C_{60} with an enzyme active site, specifically the hydrophobic binding pocket of HIV-1 protease, during the height of Bucky mania.[16] Since then researchers have investigated inhibition of glutathione-S-transferase, microsomal cytochrome P450-dependent monooxygenases and nitric oxide reductase.[17] In our laboratory a fullerene amino acid, Baa (Figure I.5), was developed and used to synthesize series of fullerene peptides through solid phase peptide synthesis (SPPS).[18] We have since reported the potent inhibition of HIV-1 protease and carbonic anhydrase by Baa and its amino protected derivatives.[19]

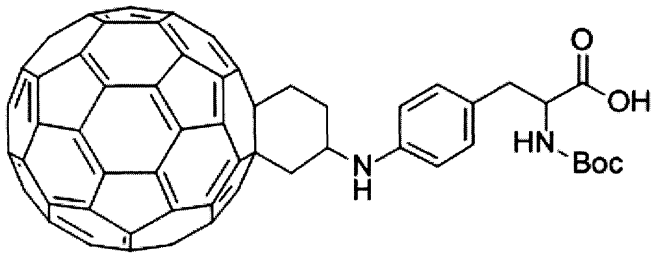


Figure I.5. The structure above is Boc-Baa, an amino protected C_{60} amino acid.[18]

Coming back to the need for improved membrane permeability for drug molecules, C_{60} and derivatives of C_{60} have been shown to be membranotropic.[20] The Baa peptides synthesized in our laboratory were shown not only to facilitate the transport of its peptide cargo, easily crossing the cell membrane, but also to cross another physical barrier, porcine skin.[21] In addition to crossing the cellular membrane, fullerene derivatives have also been shown to cross the blood brain barrier, a major challenge in the treatment of brain disorders.[22] These discoveries present a novel approach to the issue

of membrane penetration by drug therapeutics; however, since the rise of nanotechnology, there has been much concern about the biological and environmental fate of nanomaterials.[23]

There has been some debate about the toxicity of C_{60}, limiting its use in biological applications. Initial reports claiming that C_{60} was toxic to human cells and largemouth bass utilized aqueous dispersions of "nano-C_{60}" and determined that the method of sample preparation and surface derivativitization influenced the toxicological findings.[24] These early studies have since been called into question by several subsequent reports showing that the toxicity was not due to C_{60} but instead to the polar organic solvents trapped between the C_{60} crystallites.[25] In fact, one similar study utilizing aqueous C_{60} dispersions prepared with and without the use of organic solvents showed that the nano-C_{60} prepared without the aid of polar organic solvents not only showed no toxicity to rodents but actually protected the animals' liver from free radical damage.[26] Our own findings regarding the toxicity of Baa and Baa peptides to human epidermal keratinocytes (HEK cells) showed that at concentrations less than 40 $\mu g.mL^{-1}$ and 25 $\mu g.mL^{-1}$ respectively, no toxic response was observed.[27] The logical conclusion from these various reports is that any C_{60} derivative intended for biological applications should be evaluated for toxicity and the data interpreted by experienced toxicologists.

Bridging the gap between the biological and chemical worlds, nanoparticles such as C_{60} functionalized with biomolecules have the potential to revolutionize pharmacology and medicinal chemistry creating a new multidisciplinary approach to fighting disease. Chemists, inspired by the efficiency of the biological world, are arming themselves with a new arsenal of weapons to battle pandemics such as HIV. Through the intelligent design of a new class of pharmacophores, goals such as localized drug delivery and membrane penetration may be attainable.

References

[1] H. B. Vickery and C. L. A. Schmidt, *Chem. Rev.*, 1931, **9**, 169.

[2] H. E. Street and G. E. Trease, *Annals of Science*, **7**, 70.

[3] M. Kawai, H. Uneyama, and H. Miyano, *J. of Health Sci.*, 2009, **55**, 667.

[4] http://www.aloeveraibs.com/wp-content/uploads/2008/08/aminoacidstruc.jpg. Accessed 8/11/2010.

[5] R. H. Garrett and C. M. Grisham, *Biochemistry, Second Ed.*, 1999, Saunders College Publishing, Orlando, Fl., 81.

[6] http://learners.in.th/file/dawood/amino-acids.jpg. Accessed 8/11/2010.

[7] R. A. Copeland, Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists, 2005, Wiley, Hoboken, New Jersey, 2.

[8] B. Groner, *Peptides as Drug:. Discovery and Development*, 2009, Wiley-Vch Verlag GmbH & Co., Weinheim, 1.

[9] C. M. B. Edwards, M. A. Cohen, and S. R. Bloom, *Q. J. Med.*, 1999, **92**, 1.

[10] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *Adv. Drug Del. Rev.*, 1997, **23**, 3.

[11] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature*, 1985, 162.

[12] W. Krätschmer and L. D. Lamb, *Nature*, 1990, **347**, 354.

[13] A. Hirsch and M. Brettreich, *Wiley-VCH*, 2005, Weinheim, 1.

[14] T. Da Ros and M. Prato, *Chem. Commun.*, 1999, 663.

[15] (a) A. Bianco, T. Da Ros, M. Prato, and C. Toniolo, *J. Peptide Sci.*, 2001, **7**, 208. (b) D. Pantarotto, N. Tagmatarchis, A. Bianco, and M. Prato, *Mini-Rev. Med. Chem.*, 2004, **4**, 805.

[16] S. H. Friedman, D. L. DeCamp, R. P. Sijbesma, G. Srdanov, F. Wudl, and G. L. Kenyon, *J. Am. Chem. Soc.*, 1993, **115**, 6506.

[17] N. Tagmatarchis and H. Shinohara, *Mini Rev. Med. Chem.*, 2001, **1**, 339.

[18] (a) J. Yang and A. R. Barron, *Chem. Commun.*, 2004, 2884. (b) J. Yang, L. B. Alemany, J. Driver, J. D. Hartgerink, and A. R. Barron, *Chem. Eur. J.*, 2007, **13**, 2530.

[19] (a) S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, and M. G. Papadopoulos, *J. Chem. Inf. Model.*, 2009, **49**, 1139. (b) A. Innocenti, S. Durdagi, N. Doostdar, T. A. Strom, A. R. Barron, and C. T. Supuran, *Bioorg. Med. Chem.*, 2010, **18**, 2822.

[20] R. A. Kotelnikova, A. I. Kotelnikov, G. N. Bogdanov, V. S. Romanova, E. F. Kuleshova, Z. N. Parnes, and M. E. Vol'pin, *FEBS Letters*, 1996, **389**, 111.

[21] (a) J. Yang, K. Wang, J. Driver, J. Yang, and A. R. Barron, *Org. Biomol. Chem.*, 2007, **5**, 260. (b) J. G. Rouse, J. Yang, J. P. Ryman-Rasmussen, A. R. Barron, and N. A. Monteiro-Riviere, *Nano Lett.*, 2007, **7**, 155.

[22] S. Yamago, H. Tokuyama, E. Nakamura, K. Kikuchi, S. Kananishi, K. Sueki, H. Nakahara, S. Enomoto, and F. Ambe, *Chem. Biol.*, **2**, 385.

[23] (a) K. L. Dreher, *Toxicol. Sci.*, 2004, **77**, 3. (b) S. J. Klaine, P. J. J. Alvarez, G. E. Batley, T. F. Fernandes, R. D. Handy, D. Y. Lyon, S. Mahendra, M. J. McLaughlin, and J. R. Lead, *Environ. Toxicol. Chem.*, 2008, **27**, 1888.

[24] (a) S. Zhu, E. Oberdörster, M. L. Haasch, *Mar. Environ. Res.*, 2006, S5. (b) C. M. Sayes, A. A. Marchione, K. L. Reed, and D. B. Warheit, *Nano. Lett.*, 2007, **7**, 2399. (c) C. M. Sayes, A. M. Gobin, K. D. Ausman, J. Mendez, J. L. West, and V. L. Colvin, *Biomaterials.*, 2005, **26**, 7587.

[25] (a) G. Andrievsky, V. Klochkov, and L. Derevyanchenko, *Fullerenes Nanotubes Carbon Nanostruct.*, 2005, **13**, 363. (b) M. Kovochich, B. Espinasse, M. Auffan, E. M. Hotze, L. Wessel, T. Xia, A. E. Nel, and M. R. Wiesner, *Environ. Sci. Technol.*, 2009, **43**, 6378. (c) T. B. Henry, F. N. Menn, J. T. Fleming, J. Wilgus, R. N. Compton, and G. S. Sayler, *Environ. Health Perspect.*, 2007, **115**, 1059.

[26] N. Gharbi, M. Pressac, M. Hadchouel, H. Szwarc, S. R. Wilson, and F. Moussa,, *Nano Lett.*, 2005, **5**, 2578.

[27] (a) J. G. Rouse, J. Yang , A. R. Barron, and N. A. Monteiro-Riviere, *Toxicol. in Vitro*, 2006, **20**, 1313. (b) L. W. Zhang, J. Yang, A. R. Barron, and N. A. Monteiro-Riviere, *Toxicol. Lett.*, 2009, **191**, 149.

Chapter 1

Bucky-Amino Acid Acylation of pdCpA

Introduction

Marizio Prato was the first to prepare fullero-amino acids and later was the first to use the synthetic amino acids in the preparation of fullero-peptides through solid phase peptide synthesis (SPPS).[1a,b] He recognized that the novel structure of C_{60} in conjunction with an amino acid moiety may provide unique opportunities to interact with the biological world through both its shape and scale. Since that time, many others have investigated C_{60} amino acids and peptides for their membranotropic,[2-4] antioxidant,[5] and antimicrobial[6] properties. C_{60} peptides have also shown promise as enzyme inhibitors including HIV-1 protease and nitric oxide synthase.[7] Currently, fullerene peptides are prepared by solid phase peptide synthesis (SPPS); however, this technique is inherently problematic for the incorporation of fullerene amino acids into a peptide sequence.

During SPPS, hydrophobic resin beads immobilize a growing peptide chain onto the bead's surface so excess reagents and byproducts may be filtered away eliminating intermediate purification steps. Reagents are all used in large excess to drive the reaction yields to near 100%. Fullerene amino acids, with a large hydrophobic surface, irreversibly adsorb onto the bead surface (Figure 1.1). The dark color cannot be removed from the beads even after repeated cleavage reactions and many rinses. The irreversible adsorption results in less recovery of the fullerene amino acid starting material from the three to five time excess required for SPPS. Given the challenge of synthesizing and purifying C_{60} amino acids at a large scale, a more efficient procedure is needed.

In our search for a highly efficient reaction system, the most logical starting point lies in the biological world. The challenge is to find a way to exploit it. A cell produces peptides and proteins through an organelle called the ribosome. The ribosome is a large quaternary protein that reads the code within mRNA to make specific protein sequences.

The ribosome pieces together proteins based on the genetic information translated from the messenger RNA (mRNA). The mRNA contains a three base sequence called a codon that complementarily encodes for a specific amino acid carried by the transfer RNA (tRNA). In addition to its amino acid cargo, the tRNA's second important part is its anticodon (Figure 1.2 a). The anticodon is recognized through complementary hydrogen bonding of nitrogenous base pairs with the mRNA and told to drop off its amino acid cargo to become the next link in the growing peptide chain (Figure 1.2 b).[8]

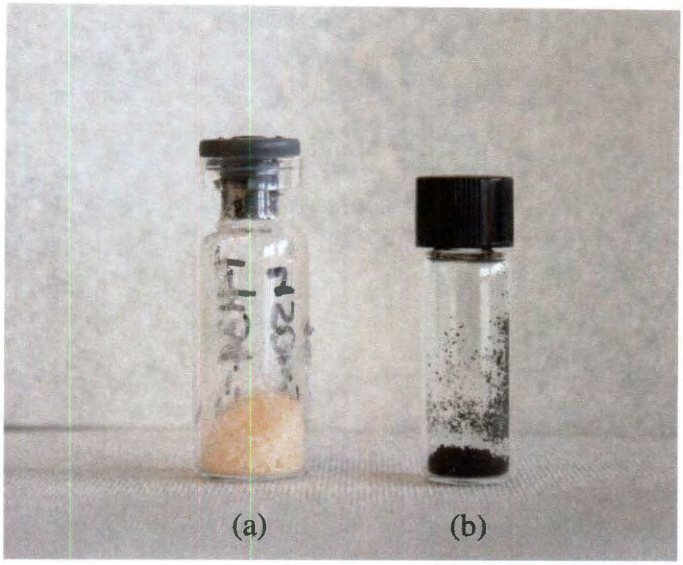


Figure 1.1. Solid phase peptide synthesis resin beads (a) before and (b) after reaction with Baa. The beads in vial (b) have been cleaved and rinsed multiple times.

There are 61 three base codons corresponding to the 20 common amino acids in addition to three nonsense or stop codons that do not correlate to any amino acid. Instead, the nonsense or stop codons tell the ribosome to terminate the growing peptide chain.[10] Eventually researchers realized they could hijack the ribosome and essentially trick it into making peptides with unnatural amino acids through the use of these nonsense codons. The field became known as non-sense suppression referring to the use of nonsense codons and suppressor tRNAs, which are responsible for recognizing nonsense codons.[10]

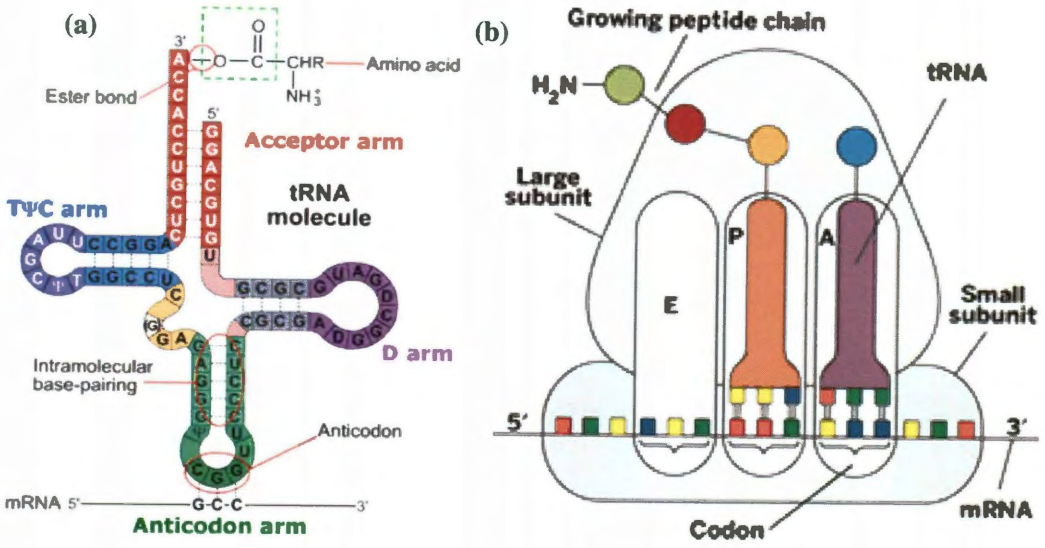


Figure 1.2. The tRNA molecule showing (a) the amino acid cargo and anti-codon and (b) the role of the ribosome, mRNA and tRNA in protein biosynthesis.[8, 9]

Hecht,[11] Chamberlin,[12] and Shultz[13] were pioneers in the work of nonsense suppression and site-directed mutagenesis. Hecht's group used a truncated suppressor tRNA, with the last two nucleotides cut off, specifically the phospho-cytidine-phospho-adenosine subunit, pCpA. A nucleotide is a subunit of RNA that includes a phosphate group, a ribose sugar ring, and a nitrogenous base, giving it its name (Figure 1.3, shows phospoadenosine or pA). The structure of DNA is made of a deoxyribose sugar ring and differs only at the 2' position where a hydrogen replaces the hydroxyl group. Hecht's methodology, utilizing the dinucleotide pCpA (Figure 1.4), was synthetically simplified by Shultz with the substitution of deoxycytidine rather than cytidine as one of the nucleoside subunits creating the hybrid dinucleotide, pdCpA, without sacrificing function (Figure 1.3).[13] The amino acylated pdCpA (pdCpA-aa), is then enzymatically ligated to the truncated tRNA and fed into the ribosome for peptide biosynthesis. Over 50 unnatural amino acids have been added to the genetic code through this technique.[10]

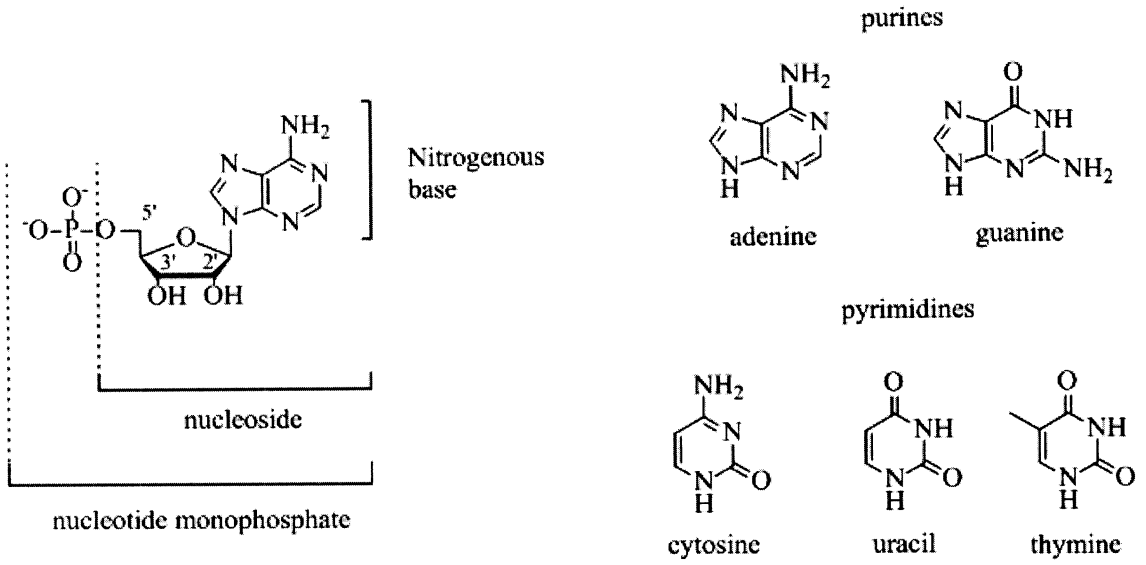


Figure 1.3. The left side shows the structure of the nucleotide phosphoadenosine (pA), the 2', 3', and 5' positions are labeled. The five nitrogenous bases found in RNA and DNA are shown on the right.

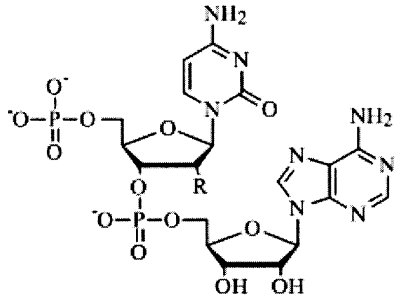
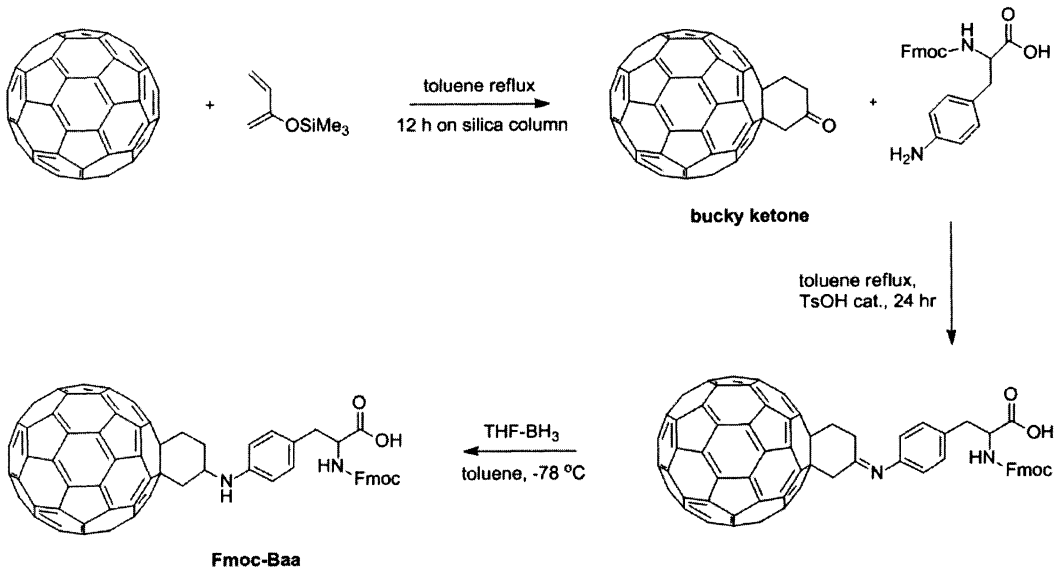


Figure 1.4. Dinucleotide structure composed of phosphates, sugar rings and nitrogenous bases, cytosine (top) and adenine. The compound is called pCpA when R = OH, and pdCpA when R = H.

Despite the laborious nature and ultimately moderate yield of synthesizing pdCpA-aa, and then enzymatically ligating it to a truncated tRNA, it remains a widely used method of incorporating unnatural amino acids through the ribosome mediated

pathway.[14] Some advancements have been made since the early days of this work to improve the yield of the final product, pdCpA-aa, and to simplify the synthesis.[15] In our ongoing effort to find efficient ways to incorporate fullerene amino acids into peptide sequences, we decided to explore this biosynthetic pathway with our previously reported Bucky amino acid (Baa).

Fullerene amino acids capable of forming peptide bonds have been prepared by many methods.[7b] Our laboratory has utilized a Diels-Alder addition of trimethylsiloxybutadiene to the fullerene cage to form a bucky ketone intermediate with a convenient handle for functionalization (Scheme 1.1). A condensation reaction with Fmoc-Phe(4-NH_2)-OH yielded an imine linkage to C_{60} which was subsequently reduced by a BH_3·THF complex. The result is an extremely stable fullerene amino acid (Baa, Scheme 1.1).[16]



Scheme 1.1. Synthesis of fullerene amino acid Fmoc-Baa through Diels-Alder addition intermediate bucky ketone.

Results and Discussion

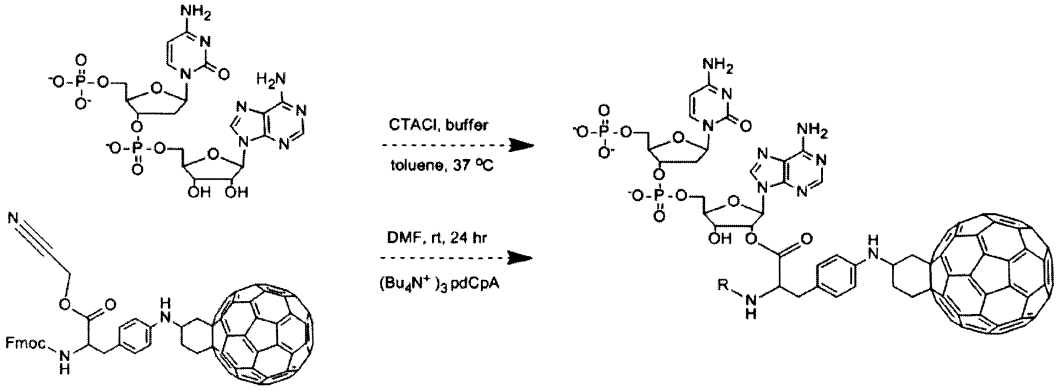
The acylation of a hybrid dinucleotide, pdCpA, with the fullerene amino acid, Baa was investigated. The unprotected hybrid dinucleotide, pdCpA (Figure 1.4), and the fullerene amino acid (Baa) were synthesized by reported literature methods.[15,16] The next and final step is amino-acylation the 2' or 3' hydroxyl of the sugar ring. Previous reports use an "activated" amino acid in the form of a cyanomethyl ester to limit nitrogen acylation of the nitrogenous bases.[13b] The challenges we are facing are two-fold: 1) pdCpA is extremely hydrophillic with three negative charges, while Baa is hydrophobic, and 2) the presence of the fullerene moiety sterically hinders the reaction altogether.

In our initial investigation two different approaches were used to bring the hydrophillic pdCpA and hydrophobic Baa into contact with each other. The first attempt utilized a cationic surfactant to exploit the inherent properties of each reactant through the formation of positively charged micelles in solution (Scheme 1.2).[17] The goal was to sequester the negatively charged pdCpA on the surface of the micelle while the hydrophobic Baa resides on the interior of the micelle. The increased local concentration of each component around and in the micelle should lead to acylation of pdCpA. No products were ever observed by this synthesis method and, in fact, in most cases the only result was hydrolysis of the cyanomethyl ester.

We next we tried the traditional approach of solubilizing pdCpA in dimethylformamide (DMF) by cloaking the negative charges with tetra-butyl ammonium (TBA) ions through ion exchange.[13] While both reactants were soluble in DMF, through this method no reaction products were observed after multiple attempts (Scheme 1.2).

In an attempt to achieve oxygen acylation in high yield, several coupling strategies, including carbodiimides and perfluorophenyl esters, were considered. In a scheme utilizing coupling reagents the free amines must be protected (phosphates are shielded as TBA salts), whereas in published methods nitrogen acylation was limited by using a cyanomethylester of the amino acid.[13] Given that Hecht has shown previously

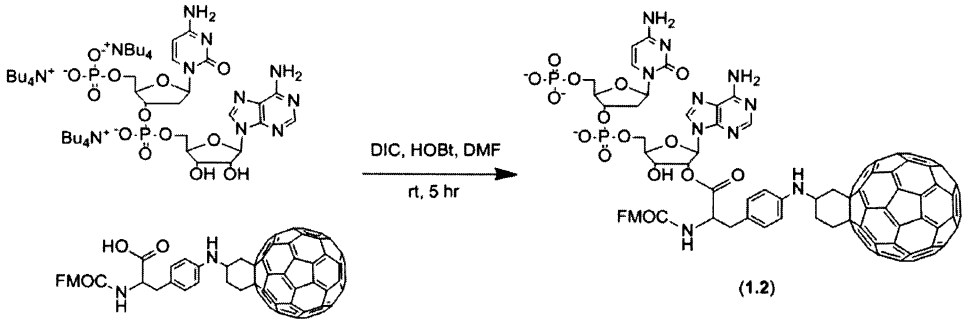
that tandemly activated suppressor tRNAs, made by bis-amino acylation of pdCpA, are capable of participating in protein synthesis, we do not need to concern ourselves with diacylation.[18] However, the use of a deoxyribose dinucleotide may be a plausible way of eliminating the possibility of di-acylation, although, as yet, no one has attempted ribosome mediated biosynthesis with an entirely deoxyribose dinucleotide. Overall, considering the steric hinderance that has prevented the acylation of the nucleotide with Baa, it is unlikely that diacylation of the diol would occur.



Scheme 1.2. Presented above are the unsuccessful coupling attempts using an activated Baa cyanomethyl ester, **1.1**, and pdCpA. The first strategy used a cationic micelle formed in solution. The second method used the TBA salt of pdCpA to enable DMF solublity. Neither method was success.

Before beginning the time consuming process of protecting already synthesized materials with the appropriate protecting groups, we needed a proof of concept that a coupling reagent would be effective in regards to the acylation of pdCpA. We utilized a diisopropylcarbodiimide (DIC) coupling reagent in conjunction with our protected amino acid, Baa, and pdCpA (Scheme 1.3). We investigated both Fmoc-Baa and the less bulky Boc-Baa through this method. After multiple attempts with DIC, some product (M + Na + H *m/z* 1815.49, calcd. *m/z* 1815.26) was detected by MALDI MS from the crude

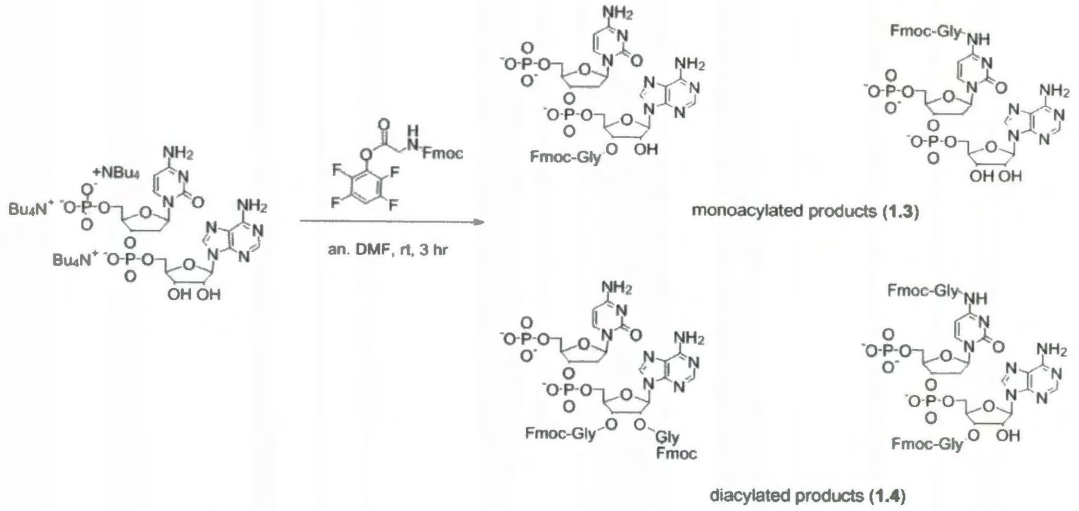
reaction mixture. However, after analysis by analytical HPLC none of the collected fractions with absorbance at 330 nm, indicating the presence of C_{60}, proved to be the correct mass to charge ratio (*m/z*) by MALDI MS. Since we were unable to isolate our compound of interest, there was no way to determine whether oxygen or nitrogen acylation had occurred. In all subsequent attempts not even acylation of the free amines was observed, which was expected at least to some degree.



Scheme 1.3. DIC coupling of Baa to the TBA salt of pdCpA to make pdCpA-Fmoc-Baa (**1.2**).

Given that all published methods and even our more aggressive coupling strategy had failed to acylate pdCpA, the reactivity of the dinucleotide diols needed to be confirmed. In this regard, a perfluorophenyl ester of Fmoc-glycine (Fmoc-Gly-OPfp) was coupled to pdCpA in anhydrous DMF (Scheme 1.4). As expected, mono- and di-acylation products were observed by HPLC and MALDI MS (Figure 1.5). To confirm we had at least some oxygen acylation, a portion of the reaction mixture was shaken with a 5% bicarb solution thereby cleaving the ester linkages but leaving amide bonds intact. The change in the chromatograms before and after the bicarb treatment confirmed that a large degree of esterfication had occurred (Figure 1.5). It is worth noting that no triacylated products were observed in the mass spectrum of the crude reaction mixture and very little diacylated product remained after the bicarb treatment. After witnessing

the ease with which the acylation with an unhindered amino acid should occur with coupling reagents, it became clear that we would have to find some other way of bringing these two molecules together.



Scheme 1.4. Coupling of hybrid dinucleotide pdCpA with Fmoc-Gly-OPfp and the possible products of the reaction. Products were characterized by MALDI MS and HPLC to determine if acylation had occured.

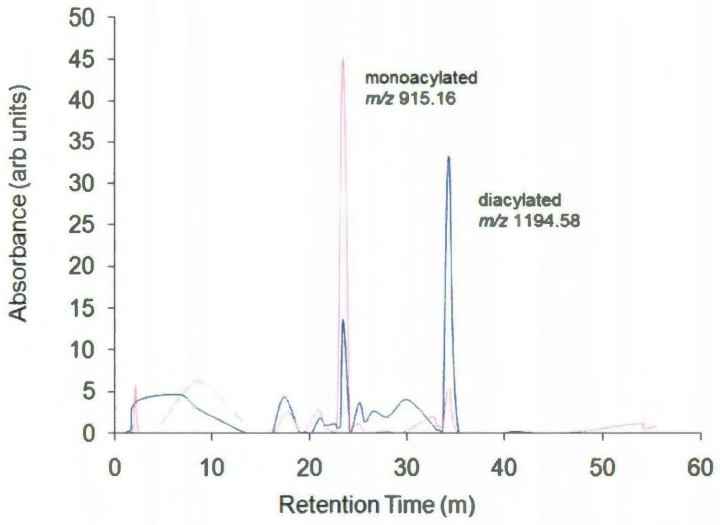


Figure 1.5. Presented above are the HPLC chromatograms (blue is before bicarb treatment and pink is after) of the reaction mixture showing major and minor products from the coupling between pdCpA and Fmoc-Gly-OPfp.

Conclusions

Our investigation into the acylation pdCpA with a fullerene amino acid led us to the conclusion that our fullerene amino acid, Baa, was entirely too hindered to react with pdCpA. Even under rigorous coupling conditions with a DIC coupling reagent, very little, if any, product was observed. However, pdCpA is easily acylated with other less hindered amino acids as we have demonstrated with a perfluorphenyl ester of glycine. Based on our results thus far we have decided to pursue alternate methods for acylation of pdCpA with fullerene amino acids. There are two possibilities: 1) to attach C_{60} as the final step after the acylation of pdCpA or 2) utilize a fullerene amino acid with a more flexible side chain such as lysine. Both of these strategies require a new route fullerene derivatized amino acids.

Experimental

Nucleoside precursors and phosphoramidte coupling reagents were purchased from Chem-Impex International, Inc., unless otherwise specified, and used without further purification. C_{60} fullerene (99.5 % purity) was purchased from MER Corp. and was purified chromatographically with toluene prior to use to remove trace oxides. All other chemicals and reagents were purchased from Sigma-Aldrich. All solvents were dried and distilled prior to use using standard techniques unless otherwise specified. Products were purified by flash column chromatography on silica gel (230 - 400 mesh).

^{1}H NMR spectra were recorded on Bruker Avance 400 and 500 MHz spectrometers. MALDI measurments were performed on Bruker Reflex or Autoflex with TOF detection. DCTB (2-[(2E)-3-(4-tert-butylphenyl)-2-methylprop-2-enylidene]malononitrile) and elemental sulfur were used as the matrices to characterize fullerene derivatives by MALDI/TOF (negative mode, $M^{\cdot -}$ ionization). α-Cyano-4-hydroxycinnamic acid (CHCA) was used as the MALDI matrix for all other

compounds. UV-Vis spectra were recorded on a Varian Cary 5000 spectrometer. High performance liquid chromatography (HPLC) was performed on a Varian ProStar instrument using a Varian C_{18} column (150 x 4.6 mm) for reaction monitoring and a Varian Dynamax C_{18} column (250 x 21.6 mm) for purification.

The fullerene amino acid, Baa,[16] and the hybrid dinucleotide, pdCpA,[15] were synthesized according to literature methods. The phosphates of pdCpA were shielded through ion exchange with tetra-butyl ammonium (TBA) ions. The resin was prepared by packing a column with a cationic exchange resin (Dowex 50Wx8, 200 mesh) and flushing it with tetra-butylammonium hydroxide (TBA^+ ^-OH, 1%, ~ 250 mL) until the eluent was basic. Next the column was flushed with deionized water until the eluent was neutral again. The nucleotide was stirred with the exchanged resin as an aqueous mixture (~ 5 mL resin/100 mg pdCpA, 5 – 10 min) in a round bottom flask in an ice bath, after which the resin was filtered out. The ion exchanged nucleotide [$(Bu_4N)_3$·pdCpA] was lyophilized to dryness and stored as a fluffy, white powder. Stock nucleotide solutions (50 - 60 mM) were prepared in anhydrous dimethylformamide (an. DMF) as needed. ^1H NMR showed a TBA to pdCpA ratio of 3.3 to 1. Concentrations of the stock solutions were determined by UV (ε_{260nm} 23,000 $cm^{-1}M^{-1}$).

Fmoc-Baa-OCH_2CN (1.1). Fmoc-Baa (125 mg, 0.1062 mmols) and triethylamine (TEA, 29.6 μL, 0.2125 mmols) were dissolved in DMF (50 mL). The Baa solution was added by syringe to an evacuated 100 mL round bottom flask fit with a stir bar and septa. Chloroacetonitrile (Cl-CH_2-CN, 20.2 μL, 0.3189 mmols) was added to the round bottom by syringe and the reaction was stirred under an Ar atmosphere. After 24 h the reaction mixture was diluted with dichlormethane (DCM, 150 mL), washed with sodium bisulfate (0.5 M, 150 mL, 2x) and then brine (150 mL, 1x). The organic phase was then dried over sodium sulfate, filtered, and concentrated under reduced pressure. The concentrate was then applied to a column of silica gel and the cyanomethylester of

Baa was eluted with 1:1 ethyl acetate:hexane. Yield: 116.2 mg, 90%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.69 (2H, m), 7.50 (2H, m), 7.33 (2H, m), 7.25 (3H, m), 6.97 [2H, d, *J*(H-H) = 7.3 Hz), 6.73 [2H, d, *J*(H-H) = 7.76 Hz], 5.15 (1H, s), 4.85 (1H, m), 4.76 [2H, d, *J*(H-H) = 15.50 Hz), 4.66 [2H, d, *J*(H-H) = 15.50 Hz), 4.40 (1H, m), 4.28 (1H, m), 4.13 (1H, m), 3.73 (2H, m), 3.54 (3H, m), 3.28 (1H, m), 3.02 (2H, m), 2.52 (1H, m). MALDI *m/z* (calcd.): $(M^{\cdot})^-$ 1215.26, (1215.22).

Carbodiimide coupling of pdCpA and Fmoc-Baa (1.2). To a 1 mL conical vial with a spin vane was added 1-hyroxybenzotriazole hydrate (HOBt hydrate, 1.1 mg, 0.008 mmols), $(Bu_4N)_3$·pdCpA stock solution (100 μL, 55 mM, 0.0055 mmols), Fmoc-Baa stock solution (113 μL, 58 mM, 0.0066 mmols) both prepared in an. DMF, and N,N'-diisopropylcarbodiimide (DIC, 1.1 μL, 0.00715 mmols). The reaction was allowed to stir for 5 h after which 2 μL of the reaction mixture was diluted to 50 μL with 1:1 acetonitrile (MeCN):0.1% TFA and centrifuged (4400 RPM, 2 min.). Analytical HPLC [0.1% aq. TFA (100 – 0%) and acetonitrile (0 – 100%) over 50 min., detection at 260 nm and 330 nm] showed only one fraction with absorbance at both 315 and 260 nm which was found to be HOBt. No product was isolated. MALDI *m/z* (calcd.): $(M + H + Na)^-$ 1815.49, (1815.26).

pdCpA-Fmoc-Gly (1.3) and **pdCpA-(Fmoc-Gly)$_2$ (1.4)** $(Bu_4N)_3$·pdCpA (100 μL, 55 mM in an. DMF, 0.0055 mmols) was added to a perfluorophenyl ester of Fmoc-glycine solution in DMF (100 μL, 55 mM in an. DMF, 0.0055 mmols) in a conical vial fit with a spin vane. The reaction was stirred at room temperature under an Ar atmosphere for 3 h after which 2 μL of the reaction mixture was diluted to 50 μL with 1:1 acetonitrile (MeCN):0.1% TFA and centrifuged (4400 RPM, 2 min.). Analytical HPLC [0.1% aq. TFA (100 – 0%) and acetonitrile (0 – 100%) over 50 min., detection at 214 nm and 260 nm] revealed two major products and almost no starting materials remaining. Appropriate

fractions were pooled and characterized by MALDI MS using a CHCA matrix. MALDI *m/z* (calcd.): $(M + 2H)^-$ mono-acylated 915.16 (914.85), di-acylated 1194.58 (1193.94).

References

1 (a) M. Maggini, G. Scorrano, A. Bianco, C. Toniolo, R. P. Sijbesma, F. Wudl, and M. Prato, *J. Chem. Soc. Chem. Commun.*, 1994, 305. (b) M. Prato, A. Bianco, M. Maggini, G. Scorrano, C. Toniolo, and F. Wudl, *J. Org. Chem.*, 1993, **58**, 5578.

2 M. E. Vol'pin, Z. N. Parnes, and V. S. Romanova, *Russ. Chem. Bull.*, 1998, **47**, 1021.

3 J. Yang, K. Wang, J. Driver, J. Yang, A. R. Barron, *Org. Biomol. Chem.*, 2007, **5**, 260.

4 I. Andreeva, A. Petrukhina, A. Garmanova, A. Babakhin, S. Andreev, V. Romanova, P. Troshin, O. Troshina, and L. DuBuske, *Fullerenes, Nanotubes, Carbon Nanostruct.*, 2008, **16**, 89.

5 J. Yang, L. B. Alemany, J. Driver, J. D. Hartgerink, and A. R. Barron, *Chem. Eur. J.*, 2007, **13**, 2530.

6 F. Pellarini, D. Pantarotto, T. Da Ros, A. Giangaspero, A. Tossi, and M. Prato, *Org. Lett.*, 2001, **3**, 1845.

7 (a) A. Bianco, T. Da Ros, M. Prato, and C. Toniolo, *J. Peptide Sci.*, 2001, **7**, 208. (b) D. Pantarotto, N. Tagmatarchis, A. Bianco, and M. Prato, *Mini-Rev. Med. Chem.*, 2004, **4**, 805.

8 S. Borman, *C&En*, 2007, **85**, 13.

9 http://www3.interscience.wiley.com:8100/legacy/college/boyer/0471661791/structure/tRNA/trna_intro.htm. Accessed 7/25/2010.

10 A. Strømgaard, A. A. Jensen, and K. Strømgaard, *ChemBioChem*, 2004, **5**, 909.

[11] (a) S. M. Hecht, B. L. Alford,Y. Kuroda, and S. Kitano, *J. Biol. Chem.*, 1978, **253**, 4517. (b) T. G. Heckler, Y. Zama, T. Naka, and S. M. Hecht, *J. Biol. Chem.*, 1983, **258**, 4492. (c) T. G. Heckler, L.-H. Chang, Y. Zama, T. Naka, M. S. Chorghade, and S. M. Hecht, *Biochemistry*, 1984, **23**, 1468. (d) J. R. Roesser, M. S. Chorghade, and S. M. Hecht, *Biochemistry*, 1986, **25**, 6361.

[12] J. D. Bain, C. G. Glabe, T. A. Dix, and A. R. Chamberlin, *J. Am. Chem. Soc.*, 1989, **111**, 8013.

[13] (a) C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, *Science*, 1989, **244**, 182. (b) S. A. Robertson, J. A. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 2722.

[14] (a) S. Ye, A. A. Berger, D. Petzold, O. Reimann, M. Benjamin, and B. Koksch, *Beilstein J. Org. Chem.*, 2010, **6**. (b) N. Muranaka, M. Miura, H. Taira, and T. Hohsaka, *ChemBioChem*, 2007, **8**, 1650. (c) N. E. Fahmi, S. Golovine, B. Wang, and S. M. Hecht, *Carbohydr. Res.*, 2001, **330**, 149. (d) T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 34.

[15] X.-F. Zhu and A. I. Scott, *Nucleosides Nucleotides Nucleic Acids*, 2001, **20**, 197.

[16] J. Yang and A. R. Barron, *Chem. Commun.*, 2004, 2884.

[17] K. Ninomiya, T. Kurita, T. Hohsaka, and M. Sisido, *Chem. Commun.*, 2003, 2242.

[18] (a) B. X. Wang, J. Zhou, M. Lodder, R. D. Anderson III, and S. M Hecht, *J. Biol. Chem.*, 2006, **281**, 13865. (b) M. Duca, S. X. Chen, and S. M. Hecht, *Org. Biomol. Chem.*, 2008, **6**, 3292. (c) M. Duca, S. X. Chen, and S. M. Hecht, *Methods*, 2008, **44**, 87. (d) D. J. Maloney, N. Ghanem, J. Zhou, and S. M. Hecht, *Org. Biomol. Chem.*, 2007, **5**, 3135.

Chapter 2

C_{60} Amino Acids through Dipolar Addition

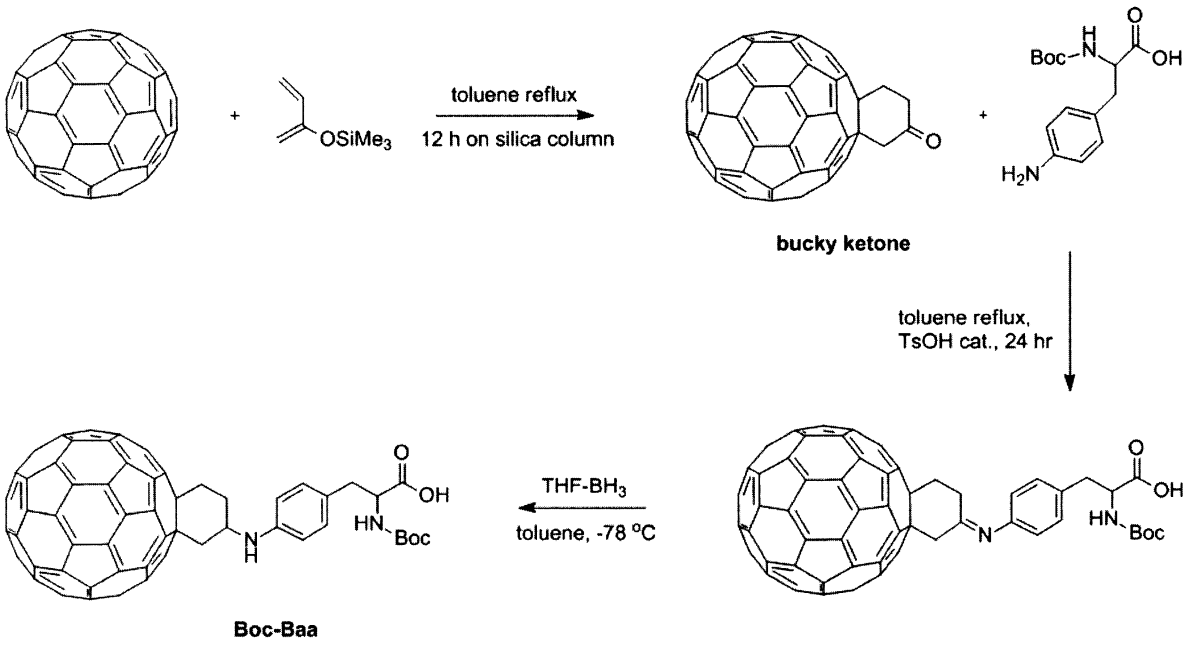
Introduction

Whether for its anti-oxidant, anti-HIV, anti-cancer, or membranotropic properties, fullerenes, specifically water soluble fullerene derivatives, have captured the imagination of researchers since their discovery in 1985.[1] As an approximately 1 nm perfect carbon sphere, C_{60}'s size, shape, and hydrophobicity afford it the unique opportunity to interact with the biomolecular world. It was inevitable that fullerenes would eventually find themselves incorporated into biologically relevant molecules such as lipids, saccharides, and oligonucleotides.[2] But perhaps the most ubiquitous use of fullerenes in the biochemical world is as an amino acid, the most basic building block of life.

Our reasons for developing a new fullerene amino acid are three-fold. First, we need a way to attach C_{60} as the final step to eliminate the steric hinderance preventing the acylation of our hybrid dinucleotide, pdCpA, as discussed in Chapter 1. Second, ideally the method is general enough for the synthesis of a variety of fullerene amino acids, both alkyl and aryl. In this regard, the yield of our flexible fullero-Lys derivative must be improved to move forward with future applications in HIV-1 PR and carbonic anhydrase inhibition.[3] Lastly, the synthesis needs to be a fast, good yield method occurring under very mild conditions.

A variety of synthesis methods exist to make fullerene amino acids; almost all utilize cycloadditions to the fullerene core. The first fullerene derivative with a convenient handle for "peptide functionalization" was prepared by Prato in 1993 through carbene addition, generated *in situ* with diazomethane, to the fullerene cage. This paved the way for other methods utilizing diazocompounds to generate carbenes.[4] Another method for producing a useful fullerene handle is the Diels-Alder addition of a butadiene

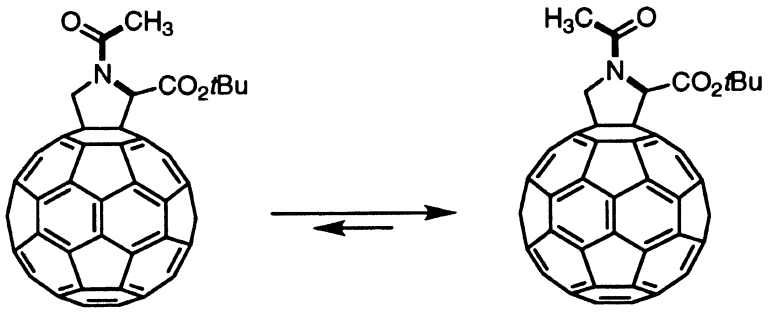
derivative to fullerene creating a Bucky-ketone intermediate.[5] The original report then reduced the ketone to a hydroxyl and used a dicyclohexyl carbodiimide (DCC) coupling to ligate the functionalized fullerene to a peptide. A similar route was reported by our lab utilizing the Bucky-ketone intermediate in a condensation reaction with an amino derivatized natural amino acid then reduced with a BH_3·THF complex, Scheme 2.1.[6] This method produces very stable fullerene amino acids in good yield (~50%); however, it suffers from long preparation times, about two weeks, even in experienced hands.



Scheme 2.1. Original Bucky-amino acid (Baa) synthesis including the Bucky ketone intermediate.

Perhaps the most widely used method of making fullerene amino acids is the addition of azomethine ylides to the fullerene cage, also called the Prato reaction. Initially, Prato's method produced a fulleropyrolidine derivative which was subsequently condensed with a protected glutamic acid residue.[7] The original method is a multistep synthesis producing several intermediates, which must be purified, before reaching the

ultimate goal, a fullerene derivatized amino acid able to be incorporated to a peptide sequence through solid phase peptide synthesis (SPPS). Later Prato improved the synthesis to create fulleroproline, Fpr, the largest non-natural amino acid to date, and the first example of a C_{60}-substituted fulleryl-amino acid.[8] However, the synthesis yields a racemic mixture of *cis-trans* isomers which must be purified on a chiral column before use in solid phase peptide synthesis (Figure 2.1).



cis *trans*

Figure 2.1. *Cis–trans* equilibrium about the tertiary amide.[8]

The final and possibly the most efficient method for the production of fullerene amino acids is through cyclopropanation, first reported by Bingle[9] and improved to a one-pot synthesis by Camps and Hirsch.[10] The Bingle reaction is one of the most efficient with yields around 50% but utilizes DBU that our nucleotide, pdCpA, cannot tolerate. Since our ultimate goal is to find a simple, fast preparation for fullerene amino acids occurring under very mild conditions, none of these methods are ideal for our purposes.

Several different cycloaddition reactions to C_{60} have been discussed including [2+1] cyclopropanation and carbene additions, [3+2] ylide addition, and finally a [4+2] Diels-Alder addition. None of them are acceptable for our purposes due to the conditions employed, specialized purification, or the time required. We are in need of a simple, mild, high-yield method of incorporating C_{60} onto the R side chain of an amino acid as

the final step in a synthesis. For these reasons, we began to explore the addition of 1, 3-dipoles to C_{60}.

Prato showed that azides react with C_{60} in a similar manner as other 1, 3 dipoles to form aza-fulleroids and aziridino fullerenes.[11] The dipolar addition to the fullerene cage is a simple method for decorating the fullerene cage with a yield consistently around 30% for the mono-addition products. The reaction conditions can be manipulated to achieve either a 6,6-closed aziridino structure by nitrene addition or a 5,6-open aza-fulleroid structure by azide addition and N_2 extrusion (Figure 2.2).[12, 12]

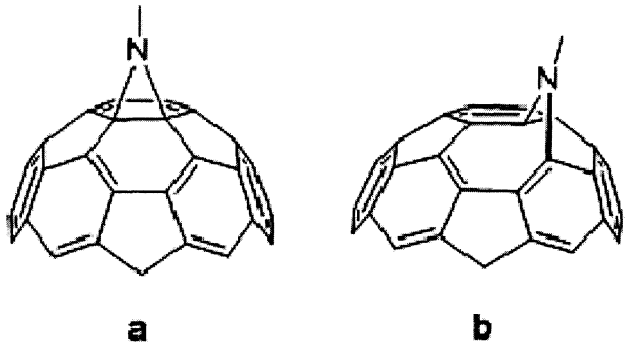
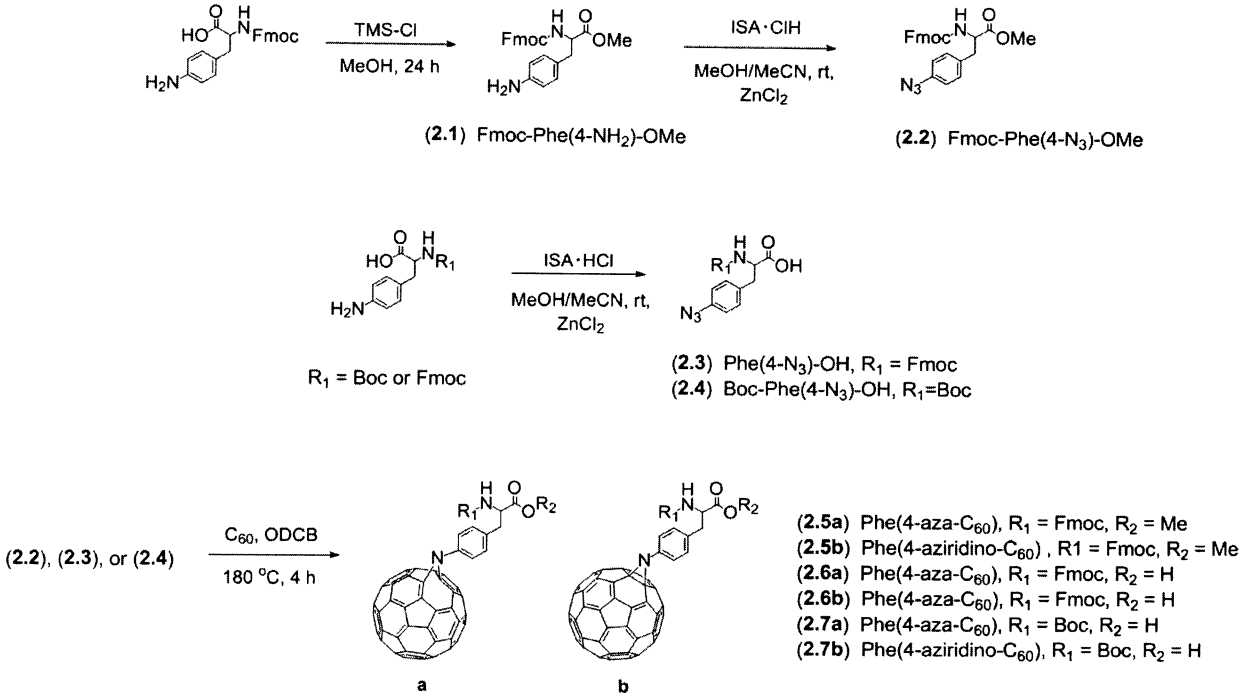


Figure 2.2. The structural unit of (a) 6,6-closed aziridino fullerene and (b) 5,6-open aza-fulleroid.

Our new approach to a fullerene derivatized amino acid can be completed in three days, including synthesis of starting materials (**2.2** - **2.4** and **2.8** - **2.10**, Scheme 2.2) and chromatographic purification, with an ~ 30% yield of the protected mono adduct, Phe(aza/aziridino-C_{60}) (**2.5** - **2.7**, Scheme 2.2). This methodology was also used successfully to prepare a protected Lys(aza/aziridino-C_{60}) derivative (**2.11** – **2.13**, Figure 2.3).



Scheme 2.2. The synthesis of phenylalanine azido starting materials and aza and aziridino fullerene amino acids.

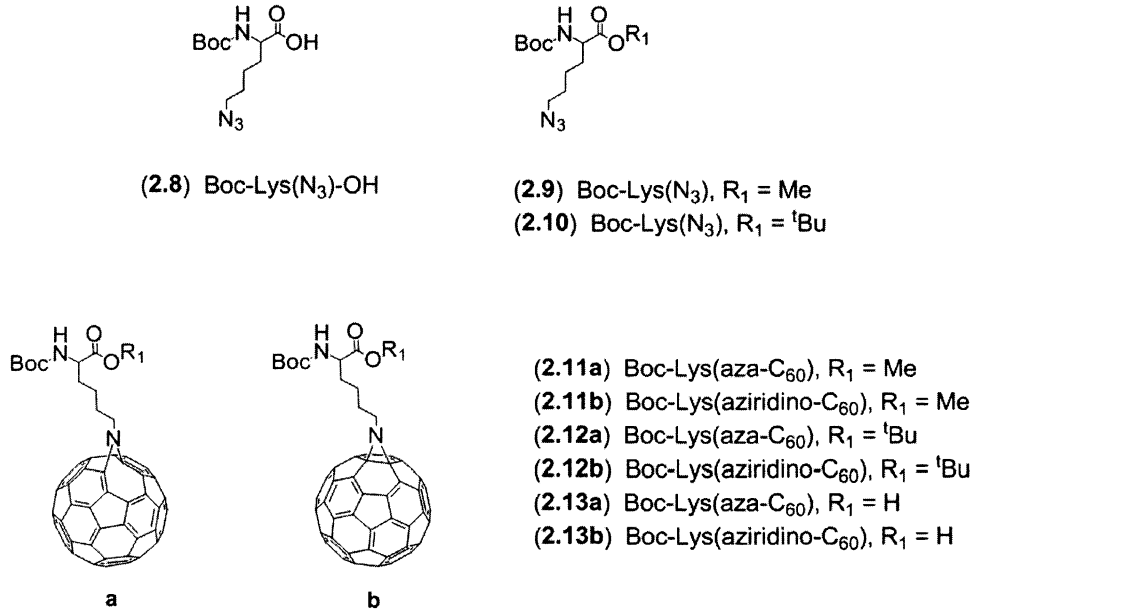


Figure 2.3. Azido-Lys starting materials and Boc-Lys(aza/aziridino-C_{60}) derivatives.

Results and Discussion

We have investigated the use of protected azido-amino acids in the synthesis of stable fullerene amino acids. At the beginning of the use of this methodology, we were aware that we would potentially generate two isomers: the 5,6-open aza-fulleroid product and the 6,6-closed aziridino fullerene product (Figure 2.2). What was not immediately obvious was which isomer might be more suited for our future applications. The kinetic aza product (5,6-open) is formed through a [3+2] azide addition to a 6,6 bond of C_{60}. A rearrangement follows in which the leaving N_2 molecule prevents the generated nitrene from re-adding at the 6,6 position. Instead the nitrene adds to a pentagon hexagon single bond generating the 5,6-open aza isomer. Under a different set of conditions the aziridino isomer is formed when the nitrene is generated before the addition to the fullerene core by thermolysis, photolysis or electrolysis.[13] The nitrene adds to a hexagon-hexagon double bond forming a 6,6-closed aziridino isomer. We employed a variety of conditions in an effort to control isomer formation, increase yields, and shorten preparation times.

Changes to the solvent system and the temperature resulted in the most significant yield increases. Initially, C_{60} was refluxed in toluene for solubility but a co-solvent was used to introduce the amino acid. C_{60} has been shown to be more easily reduced in THF solutions and when this was used as the co-solvent the yield was improved. The temperature of the refluxing toluene/C_{60} solution was consistently around 109 °C and the survival of the azido functional group was confirmed throughout the reaction by thin film IR ($\nu_{asym} \equiv 2100\ cm^{-1}$).[3] Eventually, it was discovered that both reactants were readily soluble in *ortho*-dichlorobenzene (ODCB) and, in this solvent, the reaction was complete within 24 hours (at 100 °C) with less formation of the multiple addition products.

In an attempt to push the reaction products more to the 6,6-closed isomer, the reaction was performed at 180 °C (refluxing ODCB) which should be sufficient to decompose the azide to a nitrene. Under these conditions the reaction was usually

complete within four hours (or until the disappearance of the azido amino acid by TLC). Interestingly, under these conditions, the aziridino product was still not formed to an appreciable degree and the total yield of the mono product was consistently around 30% regardless of the temperature. The thermodynamic aziridino product was isolated only when the reaction was performed at low temperature for more than 24 hours. However, allowing the thermodynamic product to form meant the formation of many more multiple addition products in the course of the reaction. Further attempts to generate the closed isomer were limited to post-purification experiments and will be discussed later.

Table 2.1. Reaction conditions for C_{60} phenylalanine derivatives.

Reaction time (hrs)	Reaction temperature (^{o}C)	Solvents	R_1, R_2	Isolated yield, open (%)	Isolated yield, closed (%)
120	110	toluene/THF	Fmoc, OH	14	-
72	110	toluene/THF	Fmoc, OH	20	13
24	110	toluene/THF	Boc, OH	19	8
24	110	ODCB	Fmoc, OH	25	2
4	180	ODCB	Fmoc, OH	32	-

Table 2.1 shows the reaction variables including reaction time, temperature, and solvent composition and includes the isolated yield of the mono open and closed products for the phenylalanine derivatives. In addition to the mono fullerene adduct, multiple addition products (up to *hexakis* adducts) are also formed as reaction time increases. Depending on the conditions of the reaction, the formation of multiple addition products may be limited. The multiple addition products were usually not isolated and may be limited by monitoring with reverse phase HPLC or adjusting the

reaction conditions. The reactions were usually stopped when *tris* products formed.

One of the goals of this synthesis is to create a general method for the preparation of fullerene amino acids and to improve the yield of our fullero-Lys derivatives. In this regard, the optimal reaction time and temperature were substrate dependant, either alkyl or aryl and in most cases the reaction was stopped when the *tris*-adducts began forming (monitored by HPLC). The temperature of the reactions were usually maintained around 90-100 °C and the survival of the azido functional group was confirmed by thin film IR (ν_{asym} = 2100 cm^{-1}).[10] Azido-lysine (alkyl) decomposes around 90 °C, resulting in undesirable reaction products (not isolated), but is avoidable by protection of the C-terminus. However, when the temperature was maintained around 60 °C, no C-terminus protection was required and no undesirable reactions occurred.

Many more multiple addition products were formed from the aryl azides relative to the alkyl azides (by HPLC) under the same reaction conditions and the yield for the mono adduct was consistently lower for Lys than for Phe. For Phe derivatives, increasing the reaction time beyond 24 hrs did not improve the yield of the mono-adduct and consumed a great deal more C_{60} in the form of multiple addition products. For **2.11**, **2.12**, and **2.13** the reaction proceeded much more slowly due to the temperature difference. These reactions typically went more than 72 hours before formation of the *tris* products.

The choice in Nα protecting group seemed to have little change on the yield and was dependent upon the future applications of the fullerene amino acid. However, it must be noted that, the aza-fullerene linkage is slightly acid labile so precautions must be taken to protect the linkage during acidic deprotection by lowering the concentration of trifluoroacetic acid (TFA, 25% in dry DCM) utilized. Piperidine deprotection of Fmoc protected fullerene compounds has been shown to form adducts with the fullerene cage making purification very difficult.[14] While we did not observe any formation of adducts by MALDI-ToF, the Boc approach proved more facile and is consistent with our current

SPPS strategy. As discussed previously, the azido-Lys starting materials had to be both amino and carboxy protected for the reaction to occur at temperatures over 60 °C otherwise a complicated mixture of products formed. The reaction conditions and yields of the fullero-Lys derivatives are given in Table 2.2.

Table 2.2. Reaction conditions for the synthesis of aza/aziridino C_{60} lysine derivatives.

Reaction time (hrs)	Reaction temperature (°C)	Solvents	R_1, R_2	Isolated yield (%)
72	60	Toluene/THF	Boc, OMe	13
72	100	Toluene/THF	Boc, OMe	22
72	100	Toluene/THF	Boc, OH	NR
72	100	Toluene/THF	Boc, O^{t}Bu	11
72	60	ODCB	Boc, OH	19

The pure isomers are easily separated by filtration through a large silica plug, eluting with toluene to remove un-reacted C_{60}, followed by a stepped gradient to elute each isomer. Two different color bands with similar R_fs of the same mass to charge ratio (*m/z*) as our desired product were isolated. The purification of the aza and aziridino derivatives is a significant improvement over Baa, in both time and solvent consumption. In addition, all of the remaining unreacted C_{60} (~70%) is recovered Detailed purification notes are included in the Experimental section.

Characterization of Isomers. The UV-visible absorption spectra of each fullerene amino acid in chloroform show maxima at approximately 260 and 330 nm, consistent with presence of the C_{60} moiety. Aziridino fullerenes (**2.5b – 2.7b**, and **2.11b -**

2.13b) exhibit an additional absorption around 420 nm.[15] Figure 2.4 shows the UV-visible spectra of the lysine derivatives in comparison with pristine C_{60}. In addition to the conserved peaks due to the C_{60} core, the spectra of **2.13b** shows a small peak at 423 nm (see Figure 2.4, inset), confirming the 6,6-closed aziridino-fullerene structure determination. The spectra of **2.13a** shows no similar absorption. The λ_{max} at 327 nm (ε 40,000) was used to determine the concentration of fullerene amino acids in solution. The colors of the aza derivatives are blue/purple, while the aziridino compounds are either amber (**2.5b – 2.7b**) or pink (**2.11b – 2.13b**).

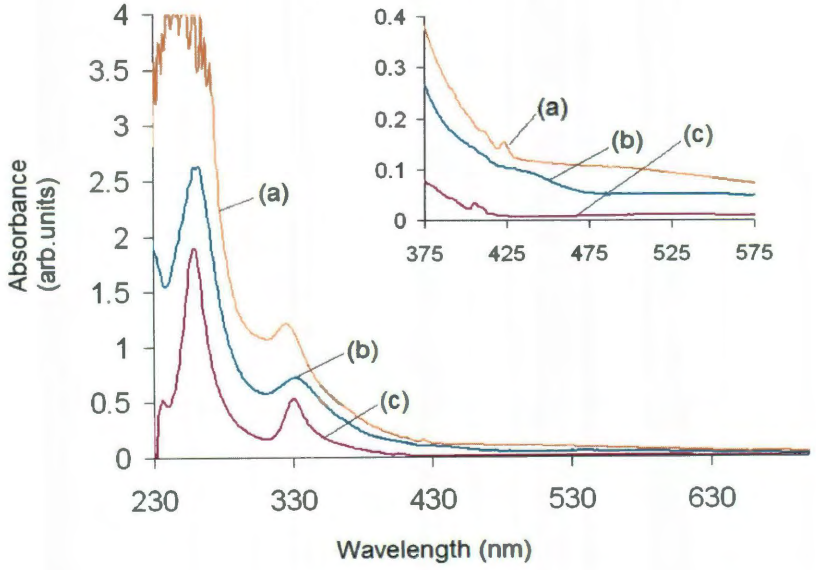


Figure 2.4. UV absorption spectra of (a) Boc-aziridino-C_{60}-Lys (**2.13b**), (b) Boc-aza-C_{60}-Lys (**2.13a**) and (c) C_{60}.

Further confirmation of the aza versus aziridino structures was obtained from the electrochemical behavior as determined by cyclic voltammetry (Figure 2.3). CV measurements were made in a mixture of DMF/toluene (3:2) at room temperature with $[Bu_4N][BF_4]$ as the electrolyte. C_{60} is known to be less easily reduced with each successive addition to its core. For C_{60} derivatives, the first electron transfer is shifted to

more negative potentials while maintaining the characteristic redox properties of pristine C_{60}.[16] In this respect, Baa (6,6-closed) is consistent with other similar closed ring structures, with its first reduction occurring at -0.377 V (Table 2.3). In contrast, in the open ring structure of the aza-C_{60} derivatives (**2.7a** and **2.13a**, Figure 2.3), only sp^2 hybridized carbons are present on the cage allowing these compounds to behave more like pristine C_{60} in their reductive ability under the specified conditions.[17] Table 2.3 shows the detailed redox potential data for the aza and aziridino-fullerene derivatives relative to C_{60} and the previously reported Baa.[18] For C_{60} and the aza-fulleroid derivatives the first reduction occurs between -0.248 and -0.288 V. However, the closed ring structures with two sp^3 C_{60} carbons do not undergo their first reduction until more negative potentials, cathodically shifted approximately 0.04 - 0.15 V relative to the aza-fulleroid derivatives.

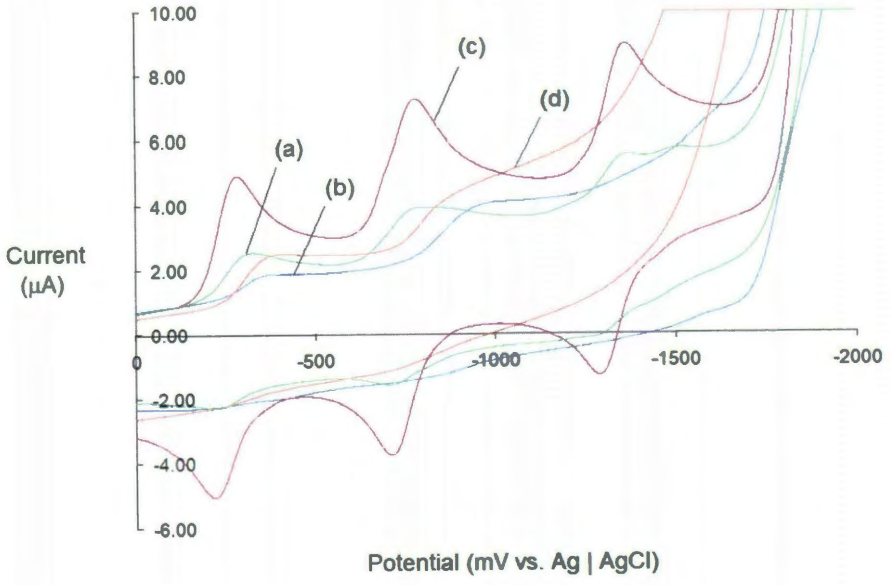


Figure 2.5. Cyclic voltammograms of (a) Boc-aza-C_{60}-Phe (**2.7a**), (b) Boc-aza-C60-Lys (**2.13a**), (c) C_{60}, and (d) Boc-aziridino-C_{60}-Lys (**2.13b**) in 0.1 M [TBA][BF_4] in DMF/toluene (3:2) at room temperature. Scan rate, 100 $mV.s^{-1}$.

Table 2.3. Selected cyclic voltammetry data for fullerene amino acids.[a]

Compound	Formal potential[a], $E_{1/2}$				Structure
	I	II	III	IV	
C_{60}	-0.247	-0.741	-1.327	[c]	-
Boc-Baa	-0.377	-0.847	-1.489	[c]	6,6-closed
Boc-Lys(aziridino-C_{60}) (**2.13b**)	-0.319	-0.792	[b]	[c]	6,6-closed
Boc-Lys(aza-C_{60}) (**2.13a**)	-0.280	-0.892	[b]	[c]	5,6-open
Boc-Phe(4- aza-C_{60}) (**2.7a**)	-0.270	-0.744	-1.318	-1.458	5,6-open

[a] Formal potentials are calculated as averages of oxidation and reduction peak potentials. Potentials are given in volts vs. a reference electrode, Ag|AgCl. [b] Third or [c] fourth reduction not observed.

Closer inspection of the voltammogram data from the 5,6-open derivative (green line) revealed a fourth reduction peak that did not fit with the rest of the pattern (Table 2.3). The first three reductions were spaced by an average of 0.55 V, whereas the fourth reduction occurs at -1.458 V, only 0.14 V from the third reduction peak. Previous reports have shown that methano-fulleroids can be electrochemically converted to the 6,6-closed isomer with a voltage of approximately -1.5 V.[19] Based on this knowledge, we surmise that the 4th reduction peak actually represented the reduction of the 6,6-closed isomer that was generated *in-situ* and therefore we may be able to utilize this electrochemical isomer conversion.

Isomer stability and conversion. It was not clear which isomer would be more stable or better suited for our applications. In addition, we need to determine whether the substituent adjacent to the aza/aziridino nitrogen (i.e. either aryl or alkyl) has an effect on the structure's stability. Fullerene and its derivatives are particularly sensitized to photo-

oxidation and degradation [20,21,22] In this regard our aza-Phe derivatives (**2.5a – 2.7a**) decomposed to C_{60} and protected Phe(4-NH_2) after several days with exposure to light and air. In the absence of light, no oxidation occurred indicating that the oxidation is photo-initiated. Over time, even under vacuum and in the dark, the aza-Phe derivatives slowly decompose.

In contrast to the aza-Phe compounds, the aza-Lys derivative, in light and air, does not decompose readily, however oxidizes to a product 32 mass units higher than the M^- ion. The acute oxygen sensitivity and subsequent ketolactam formation (+32 mass units) of alkyl aza-fulleroids is well documented in the literature.[23] Figure 2.6 illustrates the structure of the ketolactam formed by the oxidation of an aza-C_{60} derivative. No such oxidation was observed for any of our aziridino (6,6-closed) products in analogous time periods. While the study of the aza-fulleroids was not abandoned, the oxidative stability of the aziridino isomers makes them more desirable as a pharmacophore. The closed product was also the more challenging of the two isomers synthetically. As a result, we focused our efforts on converting the aza-fulleroids to aziridino fullerenes through post synthesis methods such as electrolysis.

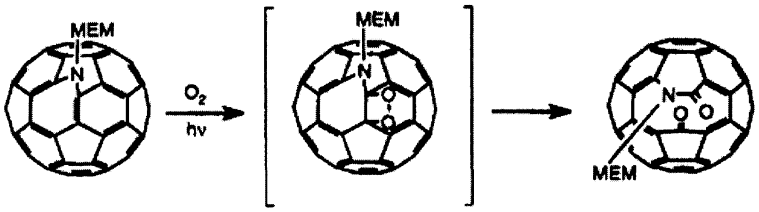
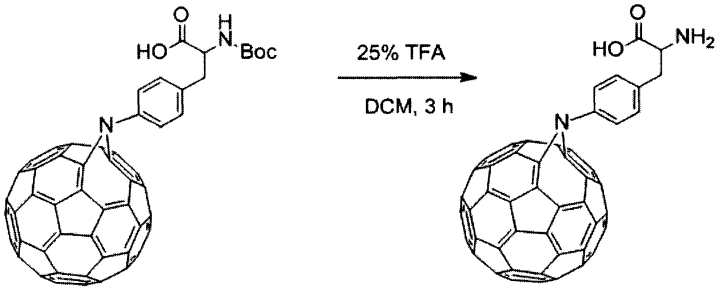


Figure 2.6. Ketolactam formation as a result of oxidation of aza-C_{60} derivatives.[23]

Several groups have demonstrated the conversion of the open product to the thermodynamically more stable closed product through photo-, thermo-, and electrochemical means.[24] All efforts to convert the kinetic aza products to the thermodynamic aziridino products thermally, as a dry powder and in refluxing toluene or

ODCB (thoroughly dried and deoxygenated), failed. Refluxing either the Phe or Lys aza derivatives in ODCB caused the formation of an uncharacterized high mass, insoluble, brown precipitate after less than 24 hours. For both derivatives, thermal conversion by refluxing in toluene (120 h.) or as a dry powder (5 h. under an Ar atmosphere) had no effect. However, the fact that the closed products had been isolated from reactions that were allowed to proceed for more than 24 hours indicated that thermal conversion is a viable method. Further attempts to convert either derivative by electrolysis (*ca.* -1.5 V) and photolysis (300 nm) resulted only in oxidation and/or decomposition.

Potential Toxicity. Given that the Phe(4-aza-C_{60}) and Lys(aza-C_{60}) are easier to prepare on a large scale, and therefore of much greater utility for subsequent studies as compared to Baa, it is important to ensure that the change in linkage unit does not alter the cellular toxicity significantly. The Boc-aza-Phe derivative was deprotected using 25% TFA as shown in Scheme 2.3. The toxic response of the deprotected Phe(aza-C_{60}) (**2.14**) was determined by MTT assay. Three different neuroblastoma cell lines (JF, SK-N-AS, and LAN-1) were used in the assay and compound **2.14** was used at concentrations from 0.004 - 0.4 $mgmL^{-1}$. No appreciable toxicity was observed for **2.14** at any of the tested concentrations after a 24 hour incubation period (Figure 2.7). In previous studies Baa was shown to be toxic at concentrations over 40 $\mu g.mL^{-1}$.[25]



2.7a **2.14a**

Scheme 2.3. Deprotection of Boc-Phe(4-aza-C_{60}) with trifluoroacetic acid.

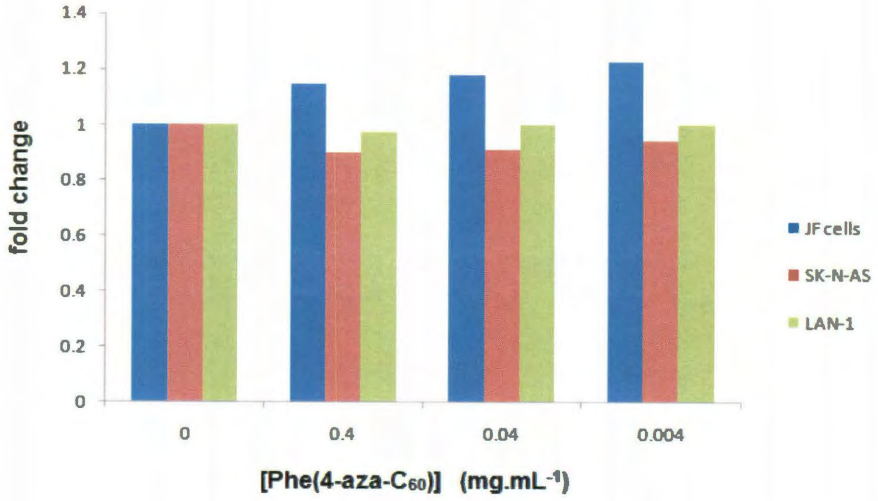
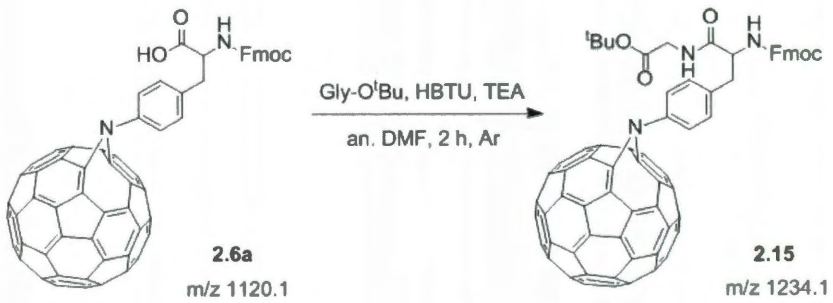


Figure 2.7. Graph depicting toxicological data obtained by incubating Phe(4-aza-C_{60}) (**2.14**) with various neuroblastoma cell lines. The toxicology experiments were performed by collaborators.

Reactivity. To demonstrate the ability of the new fullerene amino acids to be incorporated into peptide sequences, **2.6a** was coupled to tert-butyl ester of glycine (NH_2-Gly-O^{t}Bu). The reaction was carried out in 4:1 DCM/DMF in the presence of HBTU and NEt_3 in a sonication bath for 2 h (Scheme 2.4, **2.15**). The dipeptide product was detected by MALDI (*m/z* M + 1 1234.11, calcd. 1234.23) after 2 hours confirming the ability of the amino acid to form peptide bonds.



Scheme 2.4. The reaction of **2.6a** with a *tert*-butyl ester of glycine.

Conclusions

In summary, we have reported a general route to fullerene derivatized amino acids through the dipolar addition of azido amino acids, both alkyl and aryl. The ease of this procedure, including purification, is evident over previously reported methods.[6] The newly reported C_{60} amino acids differ in their linkage to C_{60}, resulting in a 5,6-open, aza-fulleroid structure as the major product (85% of isolated yield). The substituent adjacent to the aza or aziridino nitrogen affects the compound's stability to oxidation and degradation. While the aziridino isomer proved the most stable for future applications we have been unable to synthesize it in appreciable amounts and will use thermal conversion to generate the closed isomer. While the aza derivatives are acutly sensitized to O_2, the reaction does not occur in the dark and therefore they may still find some use as pharmacophores. The aza products tendancy to oxidize may be exploited in some applications, such as protein inhibition. In addition, we have demonstrated the coupling ability of the new fullerene amino acids indicating their ability for incorporation into a peptide sequence through SPPS.

Experimental

Protected amino acids were purchased from Chem-Impex International, Inc., unless otherwise specified, and used without further purification. Fullerene (99.5% purity) was purchased from MER Corp. and was purified chromatographically with toluene prior to use to remove trace C_{60} oxides. All other chemicals and reagents were purchased from Sigma-Aldrich. All solvents were dried and distilled prior to use using standard techniques unless otherwise specified. When necessary, products were purified by flash column chromatography on silica gel (230 - 400 mesh). ^{1}H NMR spectra were recorded on Bruker Avance 400 and 500 MHz spectrometers. MALDI MS measurments were performed on Bruker Reflex with TOF detection. DCTB {2-

[(2E)-3-(4-tert-butylphenyl)-2-methylprop-2-enylidene]malononitrile} and elemental sulfur were used as the matrices to characterize all fullerene derivatives by MALDI/TOF (negative mode). UV-Vis spectra were recorded on a Varian Cary 5000 spectrometer. Cyclic voltammetry data was recorded on CV-50W voltammetric analyzer.

General diazo-transfer reaction. Imidizole-sulfonyl-azide·HCl (ISA·HCl) [26] was used as the diazotransfer reagent. Initially trifyl azide (TfN_3) was investigated as the diazotransfer reagent and was prepared by literature methods.[27] The concentration of the solution was determined by assuming complete conversion of the Tf_2O starting material. Since TfN_3 is slightly explosive, it cannot be dried down completely and therefore cannot be characterized completely. The ISA·HCl was the superior reagent as it is crystalline, easily prepared on the gram scale, and has a long shelf life as opposed to trifyl azide. The subsequent diazo-transfer takes place with a divalent catalyst and a base at room temperature. The choice of catalyst and base were substrate dependant.

An amino derivatized, Nα-protected amino acid (1 eqiv.), ISA·HCl, triethylamine (2 eqiv.) and 4 mole% $ZnCl_2$ were dissolved in a solution of MeOH:MeCN (3:2) and stirred in the dark under N_2 for 12 h. The mixture was concentrated then diluted with ethyl acetate and washed with AcOH (10%, 3 x, 200 mL) and brine (1 x, 200 mL), dried over $MgSO_4$, and concentrated under vacuum. Flash chromatography is necessary at the gram scale. Reaction progress was monitored by IR for the characteristic ISA·HCl stretch (approx. 2110 cm^{-1}) to shift and the loss of NH stretching absorbencies from the starting material. The reaction was shielded from the light at all times, including work-up, to prevent decomposition of the azide.

General C_{60} amino acid synthesis. C_{60} (1 equiv.) was vacuum dried for 45 min and ortho-dichlorobenzene (ODCB) was added by cannula. To this was added the Nα-protected, azido-amino acid (1 equiv.) with stirring. The temperature was maintained between 80 - 100 °C for the Phe derivatives and 60 - 70 °C for the Lys derivatives. The reaction was monitored by RP-HPLC (5-95% IPA in 0.1% TFA over 30 min, then 95% IPA for 20 min). Typically, within 24 hrs the mono addition product was isolated by column chromatography.

Fmoc-Phe(4-NH_2)-OMe (2.1). Fmoc-Phe(4-NH_2)-OH (201.45 mg, 0.5 mmol) was added to a 250 mL of MeOH in a 500 mL Schlenk flask to form a slurry. Trimethylsilyl chloride (TMS-Cl, 126 μL, 1.0 mmol) was added slowly by syringe. All components became soluble upon the addition of TMS-Cl. The reaction was stirred under an inert atmosphere for 20 h then concentrated under vacuum. Yield: 198.8 mg, 95%. ^{1}H NMR (400 MHz, MeOD): δ 7.83 [2H, d, *J*(H-H) = 7.74 Hz, C-*H*, Phe], 7.63 [2H, d, *J*(H-H) = 8.15 Hz, C-*H,* Phe], 7.42 (4H, m, C–*H*, Fmoc), 7.32 (4H, m, C–*H*, Fmoc), 4.48 (1H, m, αC*H*), 4.29 (2H, m, CH_2, Fmoc), 4.18 (1H, m, C*H*, Fmoc), 3.75 (3H, s, CH_3), 3.62 (2H, m, αCHCH_2). MALDI/TOF, CHCA matrix, (M+1) requires *m/z* 417.17, found *m/z* 417.34.

Fmoc-Phe(4-N_3)-OMe (2.2). Compound **2.1** (41.7 mg, 0.1 mmols), $ZnCl_2$ (33 uL of 0.03 M MeCN soln., 0.001 mmol), and Et_3N (27.8 μL , 0.2 mmol) were dissolved in MeCN:MeOH (4:1, 5 mL) in a 25 mL round bottom flask and cooled in an ice bath. TfN_3[a] (214 μL, 0.12 mmol) was added dropwise to the reaction mixture in MeCN. The reaction was removed from the ice bath and allowed to slowly warm up to room temperature. The reaction was stirred in the dark for 24 h. after which it was concentrated under vacuum. The residue was diluted with Et_2O (50 mL) and filtered to remove

precipitates. The filtrate was dried under vacuum and the precipitate was washed with hexanes (3 x15 mL). Yield: 24.5 mg, 55%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.84 [2H, d, *J*(H-H) = 7.76 Hz, C-*H*, Phe], 7.62 [2H, d, *J*(H-H) = 8.13 Hz, C-*H,* Phe], 7.43 (2H, m, C–*H*, Fmoc), 7.34 (6H, m, C–*H*, Fmoc), 6.12 [d, *J*(H-H) = 7.62 Hz, N*H*], 4.44 (1H, m, αC*H*), 4.29 (1H, m, CH_2, Fmoc), 4.19 (2H, m, C*H*, Fmoc), 3.67 (3H, s, CH_3), 3.19 [1H, dd, *J*(H-H) = 5.07 Hz, *J*(H-H) = 7.36 Hz, αCHCH_2], 3.00 [1H, dd, *J*(H-H) = 6.85 Hz, *J*(H-H) = 9.39 Hz, αCHCH_2].MALDI/TOF, CHCA matrix, (M+1) requires *m/z* 443.48 found *m/z* 443.62.

Fmoc-Phe(4-N_3)-OH (2.3). ISA·HCl (1.25 g, 6 mmol), Fmoc-Phe(4-NH_2)-OH (2.01 g, 5 mmol), and $ZnCl_2$ (27.3 mg, 0.2 mmol) were dissolved in MeOH:MeCN (3:2, 300 mL) in a 500 mL Schlenk flask. Next triethylamine (1.39 mL, 10 mmol) was added. The reaction was stirred in the dark at room temperature for 12 hrs and monitored by thin film IR and TLC. Upon completion, the reaction was concentrated, diluted with diethylether (250 mL and washed with AcOH (10%, 3 x 150) and brine (1 x 200), dried over $MgSO_4$, filtered and concentrated under vacuum. At the gram scale, flash chromatography was used to yield the pure product. [10 -25 % ethyl acetate (with 1% AcOH) in hexanes]. Yield: 1.71 g, 80%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.78-7.32 (8H, m, C–*H*, Fmoc), 7.09 [2H, d, *J*(H-H) = 8.1 Hz, C-*H*, Phe], 6.93 [2H, d, *J*(H-H) = 8.55 Hz, C-*H,* Phe], 5.24 [1H, d, *J*(H-H) = 8.64 Hz, N*H*], 4.64 [1H, dd, *J*(H-H)= 4.68 Hz, *J*(H-H) = 12 Hz, αC*H*], 4.47 [1H, dd, *J*(H-H) = 7.03 Hz, *J*(H-H) = 11.0 Hz, CH_2, Fmoc], 4.37, [1H, dd, *J*(H-H) = 6.85 Hz, *J*(H-H) = 10.89 Hz, CH_2, Fmoc], 4.19 [1H, t, *J*(H-H) = 6.81 Hz, C*H*, Fmoc], 3.18 [1H, dd, *J*(H–H) = 5.39 Hz, *J*(H-H) = 14.24 Hz, αCHCH_2], 3.06 [?H, dd, *J*(H–H) = 6.07 Hz, *J*(H-H) = 14.29 Hz, αCHCH_2]. IR ν_{N3} = 2113 cm^{-1}.

Boc-Phe(4-N_3)-OH (2.4). ISA·HCl (1.25 g, 6 mmol), Boc-Phe(4-NH2)-OH (1.40 g, 5 mmol), and $ZnCl_2$ (27.3 mg, 0.2 mmol) were dissolved in MeOH:MeCN (3:2, 300 mL) in a 500 mL Schlenk flask. Then triethylamine (1.39 mL, 10 mmol) was added. The reaction was stirred in the dark at room temperature for 12 hrs and monitored by thin film IR and TLC. Upon completion, the reaction was concentrated, diluted with EtO_2 (200 mL) and washed with AcOH (10%, 3 x 200 mL) and brine (1 x 200 mL), dried with $MgSO_4$, filtered and concentrated under vacuum. At the gram scale, flash chromatography was used to yield the pure product. Yield: 1.3 g, 85%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.07 [2H, d, *J*(H–H) = 8.97 Hz, C–*H*], 6.97 [2H, d, *J*(H–H) = 8.63 Hz, C–*H*], 4.93 [1H, d, *J*(H–H) = 8.63 Hz, N*H*], 4.58 [1H, dd, *J*(H–H) = 6.64 Hz, *J*(H-H) = 14.92 Hz, αC*H*], 3.18 [1H, dd, *J*(H–H) =5.12 Hz, *J*(H-H) = 13.92 Hz, CH_2], 3.05 [1H, dd, *J*(H–H) =6.41 Hz, *J*(H-H) = 13.64 Hz, CH_2], 1.42 [9H, s, C(CH_3)$_3$]. IR ν_{N3} = 2115 cm^{-1}.

Fmoc-Phe(4-N-aza/aziridino-C_{60})-OMe (2.5 a/b). In a 250 mL Schlenk flask with a stir bar, C_{60} (39.6 g, 0.055 mmol) was vacuum dried for 45 min., then 100 mL of dry toluene was added by cannula and the mixture was heated to refluxing temperatures. Next **2.1** (24.5 mg, 0.055 mmol) dissolved in MeCN (50 mL) and added to the reaction by syringe. The reaction temperature was maintained around 100 °C for 48 h. This was a proof of concept experiment and therefore the compound was not characterized by NMR. MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 1134.03, found *m/z* 1134.51.

Fmoc-Phe(4-aza/aziridino-C_{60})-OH (2.6 a/b). In a 500 mL Schlenk flask, C_{60} (1.68 g, 2.3 mmol) and **2.1** (1.0 g, 2.3 mmol) were vacuum dried for 45 min., then 350 mL of dry ODCB was added by cannula. The reaction temperature was maintained around 165 °C for 4 hours or until the disappearance of starting materials by TLC. Due to

loss of product from precipitation, no work-up was performed. The product was purified by filtering through a large SiO_2 plug followed by stepped-gradient elution with toluene, followed by 5% EtOAc in toluene, followed by 5: 0.5:94.5 EtOAc: AcOH:toluene to elute the open isomer and 5:1.5:93.5 EtOAc: AcOH:toluene to elute the closed isomer. Yield: 825 mg, 32 %. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.78-7.32 (8H, m, C–*H* , Fmoc), 7.10 [2H, d, *J*(H-H) = 7.99 Hz, C-*H* (Phe)], 6.93 [2H, d, *J*(H-H) = 7.81 Hz, C-*H* (Phe)], 5.19 [1H, d, *J*(H-H) = 7.06, N*H*], 4.66 (1H, m, αC*H*), 4.47 (1H, m, CH_2 , Fmoc), 4.39, (1H, m, CH_2 , Fmoc), 4.20 [1H, t, *J*(H-H) = 6.75, C*H* , Fmoc], 3.19 [1H, dd, *J*(H–H) = 4.93 Hz, *J*(H-H) = 13.88 Hz, αCH-CH_2-C], 3.06 [1H, dd, *J*(H–H) = 4.79 Hz, *J*(H-H) = 13.97 Hz, αCH-CH_2-C]; MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 1120.14, found *m/z* 1120.23.

Boc-Phe(aza/ziridino-C_{60})-OH (2.7 a/b). In a 500 mL Schlenk flask, C_{60} (1 g, 1.38 mmol) and **2.2** (425.39 mg, 1.38 mmol) were vacuum dried for 45 min., then 350 mL of dry ODCB was added by cannula. The reaction temperature was maintained around 80 °C and monitored by RP HPLC for the formation of the multiple addition products (usually within 24 hours). Due to loss of product from precipitation, no work-up was performed. The product was purified by filtering through a large SiO_2 plug followed by stepped-gradient elution with toluene followed by 5:95 EtOAc:toluene, followed by 2.5:1:96.5 EtOAc:AcOH:toluene to elute the product. The pure isomers may be obtained by increasing the length of the column. Yield: 371.8 mg, 27 %. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.53 [2H, d, *J*(H–H) = 8.35 Hz, C–*H*], 7.24 [2H, d, *J*(H–H) = 8.97 Hz, C–*H*], 5.02 [1H, d, *J*(H–H) = 7.47 Hz, N*H*], 4.62 (1H, m, αC*H*), 3.22 [1H, dd, *J*(H–H) =5.71 Hz, *J*(H-H) = 12.74 Hz, CH_2], 3.15 [1H, dd, *J*(H–H) = 7.91 Hz, *J*(H-H) = 13.67 Hz, CH_2], 1.42 [9H, s, C(CH_3)$_3$] MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 998.13, found *m/z*, 998.16.

Boc-Lys(N_3)-OMe (2.9). Dissolved Boc-Lys-OMe·AcOH (80.09 mg, 0.25 mmols), $ZnCl_2$ (333 uL of 0.03 M soln., 0.01 mmol), and Et_3N (69.5 μL , 0.5 mmol) in MeCN:MeOH (4:1, 50 mL) in a 100 mL round bottom flask and cooled in an ice bath. TfN_3* (930 μL, 5.56 mmol) in MeCN was added dropwise to the reaction mixture. The reaction was removed from the ice bath and allowed to slowly warm up to room temperature. The reaction was stirred in the dark for 24 h. after which it was concentrated under vacuum. The residue was diluted with EtOAc (50 mL) and filtered to remove precipitates. The filtrate was washed with $HNaCO_3$ (2x 50 mL) and brine (1 x 50 mL), dried over $MgSO_4$, and concentrated under vacuum. Yield: 56.5 mg, 65%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.13 [1H, d, *J*(H–H) = 7.18, N*H*], 3.75 (3H, s, CH_3), 3.28 [2H, t, *J*(H-H) = 6.93, CH_2-N], 3.22 (1H, m, αC*H*), 1.81 (2H, m, αCH-CH_2), 1.64 (4H, m , CH_2-CH_2), 1.45 [9H, s, C(CH_3)$_3$]. IR ν_{N3} = 2099 cm^{-1}.

Boc-Lys(N_3)-OH (2.8). Dissolved Boc-Lys-OH (123.15 mg, 0.5 mmols), $ZnCl_2$ (666.5 uL of 0.03 M soln., 0.02 mmol), and Et_3N (104.3 μL, 0.75 mmol) in MeCN:MeOH (4:1, 50 mL) in a 100 mL round bottom flask and cooled in an ice bath. TfN_3* (3.4 mL, 0.75 mmol) in MeCN was added dropwise to the reaction mixture. The reaction was removed from the ice bath and allowed to slowly warm up to room temperature. The reaction was stirred in the dark for 24 h. after which it was concentrated under vacuum and then diluted with $CHCl_3$ and filtered. The filtrate was washed with $HNaCO_3$ (2x 50 mL) and brine (1 x 50 mL), dried over $MgSO_4$, and concentrated under vacuum to a yellow oil. Flash chromatography is necessary at the gram scale. Yield: 80.1 mg, 58%. ^{1}H NMR (400 MHz,$CDCl_3$): δ 5.15 [1H, d, J(H–H) = 8.59, N*H*], 4.33 [1H, dd, J(H–H) = 6.72, J(H-H) = 12.8 Hz, αC*H*], 3.29 [2H, t, J(H–H) = 7.43 Hz, CH_2-N_3], 1.90 [1H, m, αCH-

CH_2], 1.72 [1H, m, αCH-CH_2], 1.64 [2H, m, CH_2], 1.51 [2H, m, CH_2], 1.45 [9H, s, C(CH_3)$_3$]; IR ν_{N3} = 2127 cm^{-1}.

Boc-Lys(N_3)-O^{t}Bu (2.10) Dissolved compound **2.9** (27.3 mg, 0.1 mmols), EDC·HCl (1-Ethyl-3-(3-dimethyllaminopropyl)carbodiimide hydrochloride, 14.8 mg, 0.077 mmol), tert-butanol (7.8 μL, 0.083 mmol), and a catalytic amount of dimethylaminopyridine (DMAP) in CH_2Cl_2 (50 mL) in a 100 mL round bottom flask. The reaction was stirred in the dark until no starting material remained, ~24 h., after which it was diluted with CH_2Cl_2 and washed with HCl (10%, 3 x 50 mL) and brine (1 x 100 mL), dried over $MgSO_4$, and concentrated under vacuum to a yellow oil. Yield: 27.9 mg, 85%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.14 [1H, d, *J*(H–H) = 8.84, N*H*], 4.84 (1H, m, αC*H*), 3.60 (1H, m, CH_2-N), 3.51 (1H, m, CH_2-N), 1.91 (2H, m, αCH-CH_2), 1.78 (2H, m, αCH-CH_2), 1.68 (2H, m , CH_2-CH_2), 1.62 (2H, m , CH_2-CH_2), 1.46 (9H, s, CCH_3), 1.42 [9H, s, C(CH_3)$_3$]. IR ν_{N3} = 2099 cm^{-1}.

Boc-Lys(4-aza/aziridino-C_{60})-OMe (2.11 a/b). In a 500 mL Schlenk flask, C_{60} (180 mg, 0.25 mmol) was vacuum dried for 45 min. then 200 mL of dry toluene was added by cannula. The reaction was brought to refluxing temperatures for solubility of C_{60} then allowed to cool down to 60 °C. Compound **2.8** (86.6 mg, 0.25 mmol) was dissolved in THF (3 mL) and added by syringe to the reaction. The temperature was maintained at 60 °C and monitored by RP HPLC for the formation of the multiple addition products(usually within 72 hours). This was a proof of concept reaction performed to determine if protection of the carboxy terminus was necessary and the reaction products were only characterized by MALDI MS. The reaction mixture was washed with $NaHCO_3$ (2x 200 mL) and brine (1 x 200 mL) then dried over $MgSO_4$ and concentrated under vacuum. The residue was purified by flash chromatography eluting

with hexanes:toluene (4:1) to remove excess C_{60}, followed by a gradient elution with EtOAc in toluene (2 – 20%) to co-elute the isomers. Yield: 31 mg, 13%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.11 [1H, d, *J*(H–H) = 7.84, N*H*], 4.42 (1H, m, αC*H*), 3.78 (3H, s, CH_3), 3.77 (2H, m, CH_2-N), 2.03 (2H, m, αCH-CH_2), 1.80 (4H, m , CH_2-CH_2), 1.47 [9H, s, C(CH_3)$_3$]. MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 978.12, found *m/z* 978.44.

Boc-Lys(4-aza/aziridino-C_{60})-O^{t}Bu (2.12 a/b). In a 500 mL Schlenk flask, C_{60} (180 mg, 0.25 mmol) was vacuum dried for 45 min. then 200 mL of dry toluene was added by cannula. The reaction was brought to refluxing temperatures for solubility of C_{60}. Compound **2.10** (82 mg, 0.25 mmol) was dissolved in THF (3 mL) and added by syringe to the reaction. The temperature was maintained around 100 °C and monitored by RP HPLC for the formation of the multiple addition products (usually within 72 hours). The reaction mixture was washed with $NaHCO_3$ (2 x 200 mL) and brine (1 x 200 mL) then dried over $MgSO_4$ and concentrated under vacuum. The residue was purified by flash chromatography eluting with hexanes:toluene (4:1) to remove excess C_{60}, followed by a gradient elution with EtOAc in toluene (2 – 3%) to co-elute the isomers. Yield: 28.1 mg, 11 %. MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 1020.43, found *m/z* 1020.62.

Boc-Lys(4-aza/aziridino-C_{60})-OH (2.13 a/b). In a 500 mL Schlenk flask, C_{60} (1.0 g, 1.39 mmol) and **2.8** (272.15 mg, 1.39 mmol) were vacuum dried for 45 min. then 200 mL of dry ODCB was added by cannula. The reaction temperature was maintained at 60 °C and monitored by RP HPLC for the formation of the multiple addition products(usually within 72 hours). Due to loss of product from precipitation, no work-up was performed. The product was purified by filtering through a large SiO_2 plug, followed by stepped gradient elution with toluene to remove unreacted C_{60}, then CH_2Cl_2. Next 60:39.5:0.5 CH_2Cl_2:toluene:AcOH was used to elute the aza-derivative, followed by

79:20:1 CH_2Cl_2:toluene:AcOH to elute the aziridino product. Yield: 187.5 mg, 14%. MALDI/TOF, DCTB matrix, (M^-) requires *m/z* 964.14, found *m/z* 964.06. ^{1}H NMR (500 MHz, $CDCl_3$) (**a**) δ 5.09 [1H, d, *J*(H–H) = 7.18, N*H*], 4.43 (1H, m, αC*H*), 3.81 [2H, t, *J*(H–H) = 7.20 Hz, CH_2-N], 2.06 (1H, m, CH_2), 1.89 (1H, m, CH_2), 1.82 (2H, m , CH_2), 1.57 (2H, m , CH_2), 1.48 [9H, s, C(CH_3)$_3$]; ^{1}H NMR (**b**): δ 5.09 [1H, d, *J*(H–H) = 7.76, N*H*], 4.42 (1H, m, αC*H*), 3.69 [2H, t, *J*(H–H) = 6.90 Hz, CH_2-N], 2.05 (2H, m, CH_2), 1.91 (2H, m, CH_2), 1.82 (2H, m , CH_2), 1.62 (2H, m , CH_2), 1.48 [9H, s, C(CH_3)$_3$].

This compound has a tendency to stick on a SiO_2 column. The yield can be improved by precipitating the product with hexanes, then washing the precipitate (3 x 50 mL, 4:1 hexanes:toluene) to remove C_{60}. The pellet was then dissolved in $CHCl_3$ and filtered through a SiO_2 plug. Elute with toluene then DCM. Use 79:20:1 DCM:toluene:AcOH to elute both isomers.

Phe(4-aza-C_{60})-OH (2.14). The Boc protecting group of compound **2.6a** (25 mg, 0.02 mmols) was removed by 25 % TFA in dry dichloromethane. The residue was stirred with an excess of the reagent for 6 - 8 hrs in a glass vial. Cold Et_2O was added to the vial and the product was allowed to precipitate in the freezer overnight. The precipitate was washed with Et_2O (3 x 15 mL) and centrifuged. Yield: 16.2 mg, 90%. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.28 2H, bs, NH_2, 7.67 [2H, d, *J*(H–H) = 8.54 Hz, C-*H*], 7.67 [2H, d, *J*(H–H) = 8.66 Hz, C-*H*], 4.21 1H, m, αC*H*, 3.15 [1H, dd, *J*(H-H) = 5.82 Hz, *J*(H-H) = 13.98 Hz, CH_2], 3.09 [1H, dd, *J*(H-H) = 7.05 Hz, *J*(H-H) = 13.74 Hz CH_2]. MALDI/TOF, (M^-) requires *m/z* 898.07, found *m/z* 898.02.

Phe(4-aza-C_{60})-OH (2.14). The FMOC protecting group of **2.5a**, was removed by the addition of 0.5 mL of 20% piperidine in DMF to a solution of **2.5a** (77 mg, 0.067 mmol) in DMF (0.5 mL) and briefly sonicated then allowed to stir. After 10 minutes, the solution was diluted with 5 M HCl and centrifuged at 4400 RPMs for 5 minutes. The

water soluble fullerene amino acid supernatant was then lyophilized to dryness. Yield: 36.5 mg, 60%. MALDI MS (M+1, CHCA) requires *m/z* 898.07, found 898.28.

Fmoc-Phe(4-aza-C_{60})-Gly-O^{t}Bu (2.14). 2.5 a was coupled to tert-butyl ester of glycine to determine its reactivity. The coupling was performed in anhydrous DMF with 1 eqiv. of HBTU and 2 eqiv. of triethylamine under inert conditions. After 2 hours, MALDI/TOF (CHCA matrix) showed only starting materials and product formation at *m/z* 1234.11. Calculated (M + 1): *m/z* 1234.23.

References

1 H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature*, 1985, 162.

2 (a) M. Capaccio, V. G. Gavalas, M. S. Meier, J. E. Anthony, and L. G. Bachas, *Bioconjugate Chem.*, 2005, **16**, 241. (b) A. S. Boutorine, M. Takasugi, C. Hélène, H. Tokuyama, H. Isobe, and E. Nakamura, *Angew. Chem. Internat. Edit.*, 1995, **33**, 2462. (c) L. S. Litvinova, A. V. Gribanov, M. V. Mokeev, and V. N. Zgonnik, *Russ. J. Appl. Chem.*, 2003, **77**, 438.

3 (a) S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, and M. G. Papadopoulos, *J. Chem. Inf. Model.*, 2009, **49**, 1139. (b) A. Innocenti, S. Durdagi, N. Doostdar, T. A. Strom, A. R. Barron, and C. T. Supuran, *Bioorg. Med. Chem.*, 2010, **18**, 2822.

4 M. Prato, V. Luchinni, M. Maggini, E. Simpfl, G. Scorrano, M. Eiermann, T. Suzuki, and F. Wudl, *J. Am. Chem. Soc.*, 1993, **115**, 8479.

5 Y.-Z. An, G. A. Ellis, A. L. Viado, and Y. Rubin, *J. Org. Chem.*, 1996, **60**, 6353.

6 J. Yang and A. R. Barron, *Chem. Commun.*, 2004, 2884.

[7] M. Maggini, G. Scorrano, and M. Prato, *J. Am. Chem. Soc.*, 1993, **115**, 9798.

[8] A. Bianco, T. Da Ros, M. Prato, and C. Toniolo, *J. Peptide Sci.*, 2001, **7**, 208.

[9] C. Bingel, *Chem. Ber.* 1993, **126**, 1957.

[10] A. Hirsch, I. Lamparth, and H. R. Karfunkel, *Angew. Chem. Int. Ed. Engl.*, 1994, **33**, 437.

[11] M. Prato, C. Li, and F. Wudl, *J. Am. Chem. Soc.*, 1993, **115**, 1148.

[12] M. Prato, Chan Li, and F. Wudl, *J. Am. Chem. Soc.*, 1993, **115**, 1148.

[13] (a) D. H. R. Barton and L. R. Morgan, *J. Chem. Soc.*, 1962, 622. (b) L. Horner and A. Christmann, *Angew. Chem. Internat. Edit.*, 1963, **2**, 599.

[14] D. Pantarotto, A. Bianco, F. Pellarini, A. Tossi, A. Giangaspero, I. Zelezetsky, J. Briand, and M. Prato, *J. Am. Chem. Soc.*, 2002, **124**, 12543.

[15] C.S. Foote, *Topics in Current Chemistry*, 1994, **169**, 347.

[16] C. Boudon, J. Gisselbrecht, M. Gross, A. Herrmann, M. Rüttimann, J. Crassous, F. Cardullo, L. Echegoyen, and F. Diederich, *J. Am. Chem. Soc.*, 1998, **120**, 7860.

[17] C. Boudon, J. Cisselbrecht, and M. Cross, *Helvetica Chimica Acta*, 1995, **78**, 1334.

[18] J. Yang, L. B. Alemany, J. Driver, J. D. Hartgerink, and A. R. Barron, *Chem. Eur. J.*, 2007, **13**, 2530.

[19] M. Eiermann and F. Wudl, *J. Am. Chem. Soc.*, 1994, **116**, 8364.

[20] Jin-Pei Deng and Chung-Yuan Mou, *Fullerene Sci. Technol.*, 1997, **5**, 1033.

[21] J. W. Arbogast, A. P. Darmanyan, C. S. Foote, Y. Rubin, F. N. Diederich, M. M. Alvarez, S. J. Anz, and R. L. Whetten, *J. Phys. Chem.* 1991, **95**, 11.

[22] E . M . Hotze, J. Labille, P. Alvarez, and M. R. Wiesner, *Environ. Sci. Technol.* 2008, **42**, 4175.

[23] (a) J. C. Hummelen. M. Prato, and F. Wudl, *J. Am. Chem. Soc.*, 1995, **117**, 7003. (b) S. Xiao, Y. Li, H. Fang, H. Li, H. Liu, Z. Shi, L. Jiang, and D. Zhu, *Org. Lett.*, 2002, **4**, 3063. (c) D. M. Guldi, H. Hungerbühler, and I. Carmichael, K.-D. Asmus, and M. Maggini, *J. Phys. Chem. A*, 2000, **104**, 8601.

[24] (a) T. Nakahodo, M. Okada, H. Morita, T. Yoshimura, M. O. Ishitsuka, T. Tsuchiya, Y. Maeda, H. Fujihara, T. Akasaka, X. Gao, and S. Nagase *Angew. Chem. Int. Ed.*, 2008, **47**, 1298. (b) Z. Li and P. B. Shevlin, *J. Am. Chem. Soc.*, 1997, **119**, 1149. (c) electrolysis

[25] J. G. Rouse, J. Yang , A. R. Barron, and N. A. Monteiro-Riviere, *Toxicol. in Vitro*, 2006, **20**, 1313.

[26] E. D. Goddard-Borger and R.V. Stick, *Org. Lett.* 2007, **9**, 3797.

[27] R.-B. Yan, F. Yang, Y. Wu, L.-H. Zhang, and X.-S. Ye, *Tet. Lett.*, 2005, **46**, 8993.

Chapter 3

Fullerene-Based Inhibitors of HIV-1 Protease

Introduction

As discussed in the previous chapters, fullero-peptides utilizing an unnatural C_{60} amino acid in the sequence are not a recent construct. Prato was the first to report the synthesis of a fullerene peptide, recognizing the novel properties a fullerene might impart to biological molecules.[1] Since that time our laboratory has produced several fullerene peptides to investigate their cellular penetrating properties and their inhibition of neuroblastoma, a cancer affecting children.[2] Others have investigated the cytotoxicity, DNA cleavage, and enzyme inhibition by fullerene amino acids, peptides and other C_{60} derivatives. [3] One of the enzymes studied in regards to inhibition by fullerene is human immunodeficiency virus-1 protease (HIV-1 PR).

When a person is infected with HIV the virus encodes for several enzymes, including reverse transcriptase (RT), integrase, and protease. The first two enzymes are responsible for transcribing and integrating the virus's genetic material into that of the host's, thereby hijacking the host's cells to produce viral proteins. HIV-1 PR is one of the viral proteins produced. It is an aspartic protease responsible for catalyzing the hydrolysis of precursor polyproteins associated with HIV infection.[4] Without the function of either RT or PR, the virus fails to mature and does not become infectious. Both, RT and PR have become popular targets for HIV inhibition.[5]

Aspartic proteases (Asp PRs) are a broad class of enzymes involved in many body processes, such as digestion and pregnancy. Aspartic proteases are characterized by optimal activity in the low pH range (1 – 5), two catalytic aspartic acid residues present in the dimerized protease, and inhibition by pepstatin A, a peptide inhibitor of all Asp PRs (Figure 3.1).[4] Asp PRs have been isolated from most forms of life, including plants, vertebrates, fungi, and retroviruses, where they are responsible for cleaving immature

polypeptide substrates.[6] HIV-1 PR was one of the first retroviral proteases to be studied by X-ray diffraction[7] and pepsin, a digestive enzyme and aspartic protease, was the first protein ever to be analyzed by X-ray diffraction.[8]

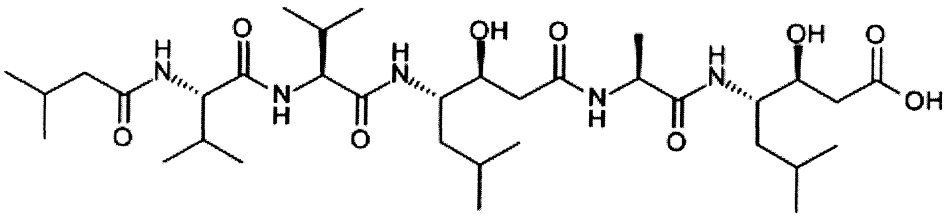


Figure 3.1. The structure of pepstatin A, a peptide inhibitor of all aspartic proteases is shown.

The structures of Asp PRs are somewhat variable with about 5% of the structure conserved between all strains.[6] Functional retroviral aspartic proteases form a homodimer, with two identical monomers each contributing a highly conserved region near the catalytic aspartic acids (Figure 3.2). The highly conserved region consists of a triad of residues, Asp-Thr-Gly, and other hydrophobic residues, such as alanine and phenylalanine.[9] In their monomeric form retroviral PRs are catalytically inactive. In the highly conserved region, the aspartyl residues are in close proximity to each other and coordinate to a water molecule; it is believed that this coordination is a major source of the stability of the dimer as a whole.[10] Depending on the presence and structure of the inhibitor, only one of the aspartyl residues is protonated at a time, due less to the pH of the environment and more to their proximity to each other.[4] In fact, the two aspartyl residues have different pKa's of 3.1 and 5.2 for the deprotonated and the protonated residue respectively.[11]

In addition to the asapartyl residues, the dimerized enzyme contains a binding pocket made up of various hydrophobic amino acids. This forms the active site, or of HIV-1 PR and is the region where mutations often occur in response to specific

interactions with inhibitors.[4] The active site is covered over by one or two binding loops or flaps, depending on whether they are cellular or retroviral PRs respectively, which take part in binding to substrates and inhibitors. For nomenclature purposes the binding pocket is divided into subsites composed almost entirely of hydrophobic residues. Figure 3.3 shows the standard nomenclature used to designate both binding pocket subsites and the corresponding interaction points on the peptide substrate in addition to the location of the scissile, or cleaved, bond.[4]

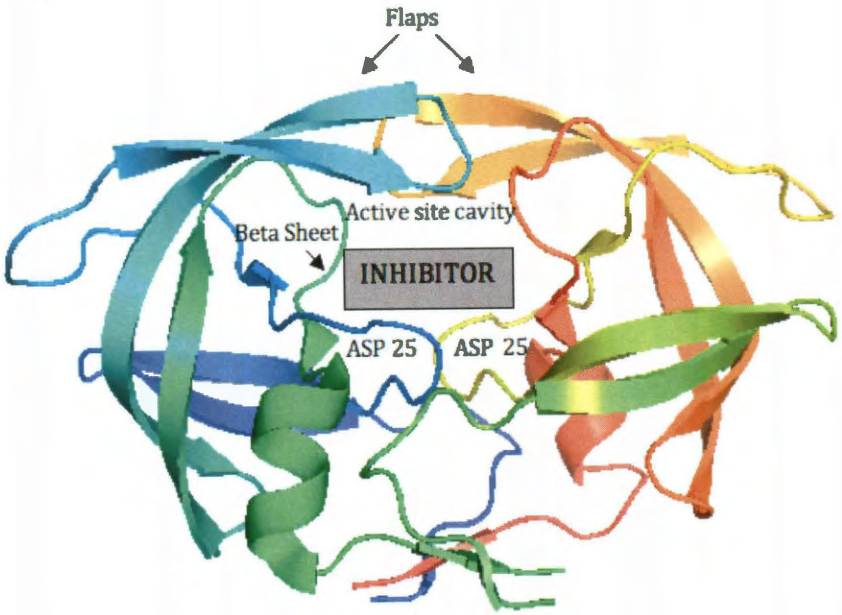


Figure 3.2. The dimerized structure of HIV-1 PR illustrating the location of the catalytic aspartyl residues, the flaps, and the binding pocket where the substrate or inhibitors fit is shown above.[12]

In a general acid–base mechanism, the nucleophilic aspartyl residue initiates peptide hydrolysis by deprotonating a coordinated water molecule, or thiol in some cases, thereby generating a nucleophile capable of cleaving the amide bond. Although the exact mechanism is not known, Figure 3.4 shows the generally accepted mechanism of

catalysis of retroviral Asp PRs whether that be through a concerted reaction or through a series of separate steps.[13]

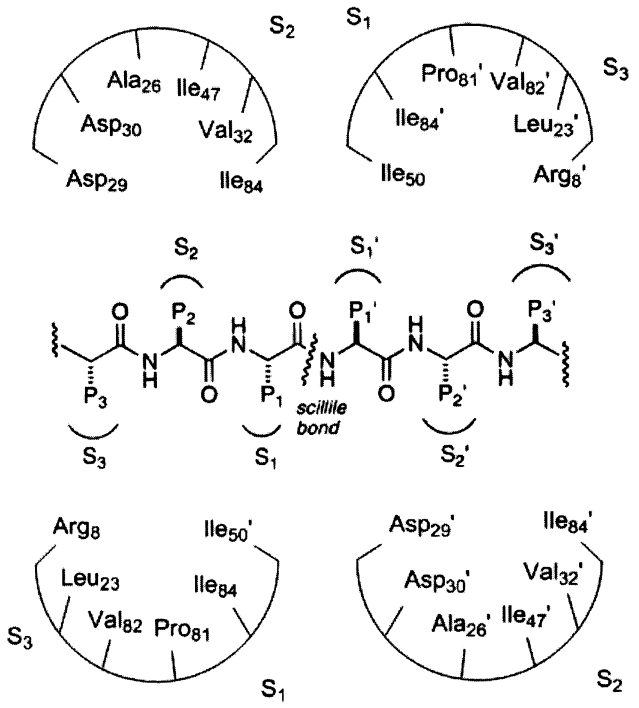


Figure 3.3. This is the standard nomenclature used to reference the peptide substrate residues (P_1-P_n) and their corresponding binding sites on HIV-1 PR (S_1-S_n), including the specific amino acids present, in relation to the scissile peptide bond. The prime notation (') refers to one monomer of the dimerized protease.[10]

The determination of the structure and potential mechanism of Asp PRs was a result of the initial sequencing and crystal structure of pepsin, the digestive enzyme, and through co-crystalization of various Asp PRs with the inhibitor pepstatin.[9-10] Those events allowed researchers to intelligently design several classes of new inhibitors based on a remarkably accurate model of the binding pocket and its relevant interactions with inhibitors. Of primary importance were: (1) displacement of the nucleophilic water

molecule, (2) hydrogen bonding between the enzyme and the peptide backbone of pepstain, (3) VDW interactions between the sidechains of pepstatin and enzyme residues, usually hydrophobic, and (4) a conformational change in the flap of the enzyme allowing more favorable hydrogen bonding opportunities with pepstatin.[10] Through the combined efforts of many researchers, several classes of aspartic protease inhibitors (PRIs) emerged.

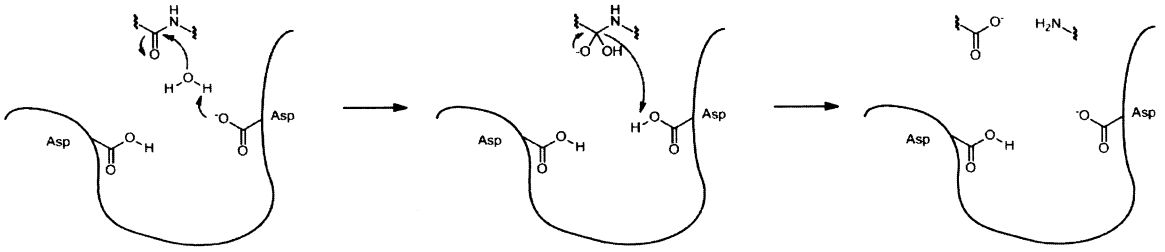


Figure 3.4. The proposed acid-base mechanism of proteolytic cleavage, including the tetrahedral intermediate, by this type of aspartic protease.[14a] The other type of protease is a metallo-protease in which a zinc ion is coordinated to the substrate as a Lewis acid and generating an activated carbonyl.[10]

Since the substrate and the first known inhibitor of HIV-1 PR, pepstatin A, are both peptides, most early inhibitors were either peptides or were designed as substrate mimics. One observation made from these early studies was that HIV-1 PR was able to recognize and cleave the Tyr-Pro or Phe-Pro (Figure 3.5) dipeptide sequences of viral polyproteins, unlike cellular proteases, and it was correctly thought that this condition may impart some level of specificity for HIV-1 PR.[14] The knowledge gained from these early inhibitors along with the knowledge that enzymes lower the activation energy of a reaction through specificity for either or both the substrate and the transition state allowed the development of another class of PRIs.

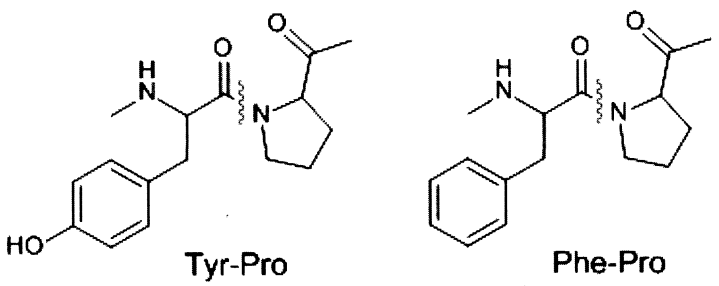


Figure 3.5. Dipeptide subunit recognized and cleaved by HIV-1 PR but not by cellular PRs. The red line represents the scissle bond that is cleaved during proteolysis.

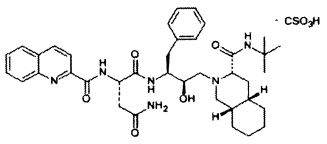
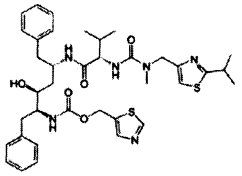
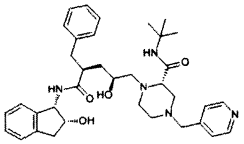
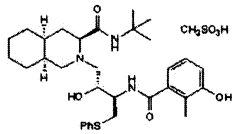
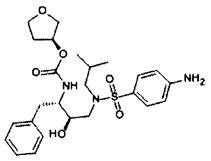
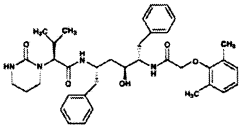
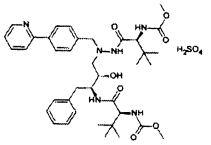
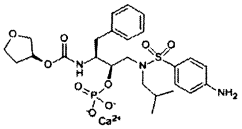
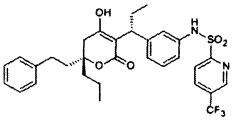
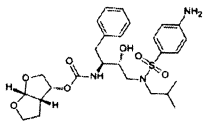
Based on the previous studies, a new series of peptidomimetic inhibitors were designed to imitate the tetrahedral transition state of the enzyme catalyzed reaction. Several structures were considered to emulate the proteolytic transition state of the Tyr-Pro or Phe-Pro dipeptide structure; the most successful of which was a nonhydrolyzable hydroxyethylamine subunit.[4] In fact, nine out of ten of the currently available, FDA approved, inhibitors use a nonhydrolyzable hydroxyethylene or hydroxyethylamine subunit at the core (Table 3.1), one as a prodrug. In addition to the specificity for the dipeptide sequence, it was also discovered that an R stereoconfiguration at the central carbon bearing the hydroxyl group was preferred for short protease inhibitors (PRIs). [15] The crystal structures of such transition state isosteres with HIV-1 PR confirm that they make favorable interactions with the two catalytic Asp residues where the coordinated water molecule usually resides.[15] There are currently ten marketed, FDA approved, protease inhibitors for HIV-1 PR; they are depicted in Table 3.1 in the order of their FDA approval, along with the stereoconfiguration at the hydroxyl bearing carbon and their inhibition constant, K_i.

The potency of an enzyme inhibitor is evaluated by monitoring the kinetics of the enzymatic reaction. This is accomplished by quantifying substrate depletion or product formation through chromogenic, radiometric, or HPLC based assays. [15] The values are usually given as EC_{50}, the dose required to effect a 50% decrease in the initial reaction

velocity, or as K_i, the inhibition constant; both are reported in units of molarity. The inhibition constant is actually the dissociation constant of the enzyme/inhibitor complex and is considered to be a more accurate reflection of the pharmacokinetics of the inhibitor with the enzyme. In many cases the EC_{50} is roughly equivalent to the apparent K_i. However, in the case of tight binding inhibitors such as ours, the EC_{50} depends on the enzyme concentration and other conditions of the assay such as temperature and buffers. An inhibitor is considered to be tight binding when the inhibitor concentration used in the assay is near the total enzyme active site concentration, $[I] \approx [E_t]$. Due to their dependence on assay conditions, EC_{50} values are not always comparable to each other whereas the inhibition constant, K_i, is a measure of the rate of dissociation of the complex and is therefore comparable to any other K_i measurement. In any case, the EC_{50} measurement is a less rigorous evaluation and is commonly used in high through put screening of inhibitors to determine lead compounds and can be converted to the K_i if other kinetic parameters of the substrate are known.[20]

The effectiveness of PR inhibitors has been diminished by the virus's high rate of mutation forcing patients to take "cocktails" of drugs to combat the constantly changing virus. There have been at least 1929 drug induced mutations isolated from "resistant" patients. Most mutations occur in the binding pocket in response to the P_1/P_3 sites on inhibitors.[4] One of the FDA inhibitors, tipranavir, is a non-peptidic PRI and is only administered in conjuction with ritonavir as a combination therapy. It is an extremely potent inhibitor that does not use the hydroxyethylamine subunit and has been effective in patients with drug resistance due to mutations. Tipranavir is only administered to patients that have shown resistance to multiple protease inhibitors as it can cause life threatening side effects such as severe liver damage.[17] To date, tipranavir combination therapy is the most successful, albeit hazardous, strategy for combating the ever changing virus. The need for structurally novel inhibitors of HIV-1 PR is evident.

Table 3.1. FDA approved protease inhibitors. [4, 16]

Generic Name	K_i (nM)	Stereoconfiguration of C-OH	Structure
saquinavir mesylate	0.12	R	
ritonavir	0.01	S	
indinavir	0.56	S	
nelfinavir mesylate	2	R	
amprenavir	0.6	R	
lopinavir and ritonavir	0.003	S	
atazanavir sulfate	0.75	S	
fosamprenavir calcium	80 (IC_{50})	R	
tipranavir for resistance	0.01	-	
darunavir	0.0045	R	

In an effort to find new and potent inhibitors of HIV-1 PR, many molecules have been considered. The fit of C_{60} with the binding pocket of HIV-1 PR was first proposed by Friedman in 1993, and in fact, was initially a joke between frustrated graduate students. The joke evolved into an epiphany as Friedman looked more closely at the ~1 nm diameter, hydrophobic binding pocket in relation to the buckyball. He realized that C_{60} would actually make quite a nice fit, both through steric and electrostatic interactions.[17] Since Friedman's initial work many others have attempted to improve the affinity of fullerene derivatives for the active site of HIV-1 PR.[18] Despite many attempts by several researchers to make a potent, water soluble, fullerene PRI, the most potent of the fullerene inhibitors (103 nM) was designed by Friedman and co-workers by modeling hydrophobic desolvation upon complex formation with PR.[19] Figure 3.6 shows the most potent fullerene inhibitors of HIV-1 PR, their inhibition constant, K_i, when available, and/or their EC_{50}.

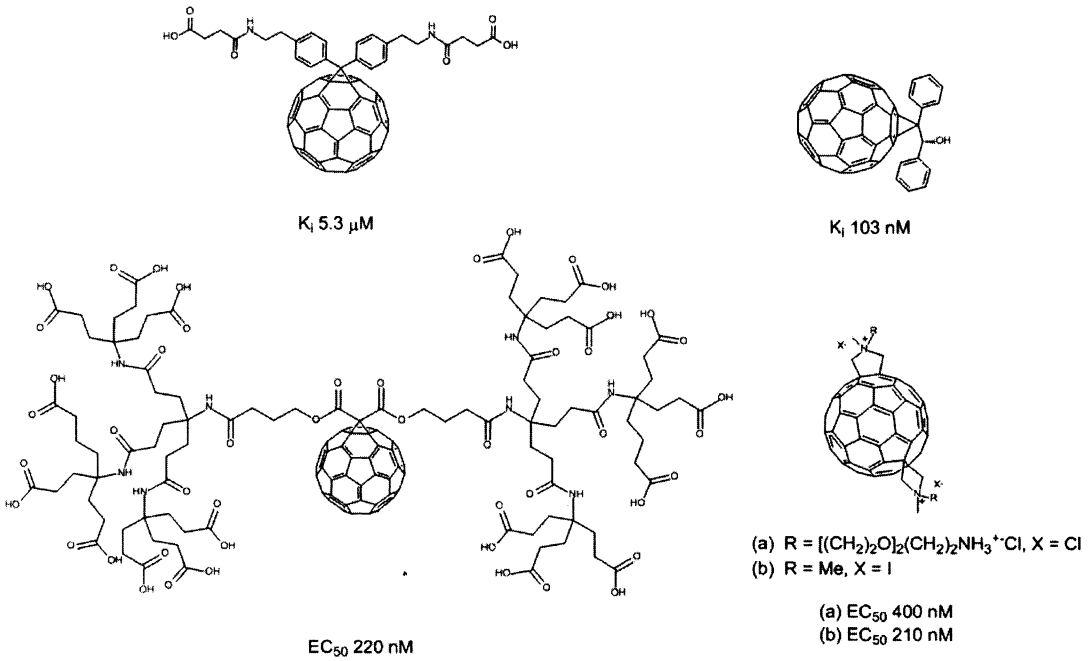


Figure 3.6. The structures of selected fullerene based PRIs are shown above with their K_i or EC_{50}.

New enzyme inhibitors are screened by evaluating the kinetics of the enzymatic reaction. The kinetic behavior of many simple enzymes is approximated by Michaelis-Menten (M-M) kinetics.[20] M-M kinetics assumes that there is no intermediate of the reaction, no cooperativity with other substrates, enzymes or cofactors, and no product inhibition during the course of the reaction. M-M kinetics requires that the rate of product formation is linear with time and proportional to both substrate and enzyme concentration. The corresponding reaction progress curve to determine initial reaction velocity, v_0, is a rectangular hyperbola following Eq. 3.1, where V_{max} is the maximum rate or velocity, [S] is the substrate concentration and K_m is the M-M constant. The K_m is defined as the substrate concentration at which the reaction velocity is at half its maximum value, $\frac{1}{2}V_{max}$. It is a kinetic constant and is meant to provide a consistent method of evaluating substrate saturation of an enzyme active site. The value of K_m depends on environmental factors such as pH, temperature, and ionic strength. In an inhibition assay the substrate concentration should be near the K_m to prevent competition with the inhibitor. [21]

$$v_0 = \frac{V_{max}[S]}{K_m + [S]} \tag{3.1}$$

The K_m of a substrate is determined by evaluating the initial velocity, v_0, of substrate hydrolysis at various substrate concentrations in the absence of an inhibitor. During the initial phase of the reaction, when substrate depletion should be < 10 - 15% of the starting value, a steady state is reached where [S], [ES], the concentration of the enzyme substrate complex, and the reaction rate are constant and the reaction velocities are well described by a plot of product formation versus time. In a typical kinetic assay, enzyme concentrations are in the nM range while substrate concentrations are in the

micro to milli molar range. This condition ensures that the substrate concentration [S] is roughly constant throughout the "initial phase" of the reaction. The kinetic parameters K_m and V_{max} can be determined by fitting the data from the reaction progress curve through non-linear regression using the M-M equation (Eq. 3.1).[20]

In addition to the K_m and V_{max}, the total enzyme active site concentration, $[E_t]$, must also be determined to evaluate the inhibition of an enzyme. The concentration of HIV-1 PR added to an assay is not the same as the total active site concentration. An enzyme is in constant flux between dimerized and monomeric forms. Even then the dimerized protein may not be in an active conformation; therefore, the $[E_t]$ must be determined experimentally. The empirical determination of $[E_t]$ is accomplished by titration with a tight binding inhibitor. As discussed previously, a tight binding inhibitor is effective at low concentrations somewhere near the total enzyme concentration.[21] In plot of fractional velocity (v_i/v_0) versus [I] the point where the linear portion of the curve intersects the *x*-axis is the total enzyme active site concentration, $[E_t]$.

The kinetic behavior of HIV-1 PR with our inhibitors was evaluated by a cell-free fluorescence resonance energy transfer (FRET) based assay using a FRET peptide substrate. FRET based assays have become a popular method for high throughput screening of potential pharmacophores.[22] FRET is a distance dependant interaction and a FRET peptide consists of three main parts: a donor fluorophore (D), a quencher or acceptor molecule (A), and a linker between the two. When the donor is excited, its energy is transferred to the acceptor without the emission of a photon. Usually the emission wavelength of the D moiety overlaps with the absorption wavelength of the acceptor. When the linker/peptide is cleaved the donor and acceptor are separated and the energy is no longer transferred resulting in a quantifiable fluorescence signal.[23] The FRET peptide used in our assay is derived from the native cleavage site on the polyprotein substrate of HIV-1 PR.[23]

In our investigation of fullerene inhibitors of HIV-1 PR we wanted to answer several questions: 1) will a fullerene amino acid be a good base for the design of fullerene-based PRIs; 2) will subtle changes to the C_{60} amino acid core affect the achieved inhibition; 3) can we improve the inhibition of the fullerene amino acid through the incorporation of a designer peptide tail; and 4) what properties of HIV-1 PR can we exploit to improve the observed inhibition? Given that the synthesis and purification of fullerene compounds can be time-consuming and expensive, the use of a molecular docking (MD) simulation aided us in the discovery of lead compounds and in the design of new inhibitors.[24] In collaboration with a group of theoretical chemists with a molecular docking (MD) simulation relating to C_{60} derivatives with HIV-1 PR [25] and a group of biologists, we set out to experimentally verify the predictive ability of their modeling program with our fullerene amino acids. With the knowledge gained from these experiments, the ultimate goal was to design more active, fullero-peptide inhibitors of HIV-1 PR.

The rational for the design of our first peptide inhibitor was based on the structure of HIV-1 PR itself. Our intention was to exploit the structure of HIV-1 PR for our purposes with a chain of positively charged amino acids which should result in the formation of a salt bridge with the aspartyl residues. Our hypothesis was that a poly-Lys peptide tail would provide both the positive charges and enough flexibility for the electrostatic interaction to occur. In addition to the electrostatic interaction, the poly-Lys tail should also improve the aqueous solubility of the C_{60} amino acid. Even without the electrostatic interactions, at a minimum, the poly-Lys tail should provide many H-bond donors and acceptors. We used Lipinski's rule of five for drug candidates as a guide in our inhibitor design.[26] The rule of five states that a drug lead should have no more than five H-bond donors, ten H-bond acceptors, a molecular weight under 500 and an octanol-water partition coefficient of 5 to be "druglike" and orally available. A tripeptide was determined to be the minimum length tail necessary to achieve our goals despite

disregarding most of Lipinski's guidelines. The structure of the tripeptide tail is shown in Figure 3.7.

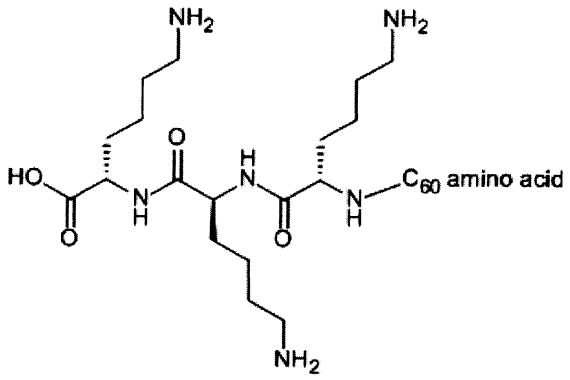


Figure 3.7. The structure of the poly-Lys peptide tail.

Ultimately we wanted to design a transition state mimic with our peptide structure. In the design of our transitions state isosteres we used the knowledge that HIV-1 PR can recognize a Phe-Pro dipeptide subunit. Utilizing our C_{60} amino acid as the Phe component, we designed a sequence including Pro-Hyp-Lys (Figure 3.8a). The proline residue provides the recognition element required for specificity to HIV-1 PR. The hydroxyproline residue provides the tetrahedral transition state analog with R stereoconfiguration at the hydroxyl bearing carbon. The final lysine residue was added both for aqueous solubility and to provide additional H-bonding opportunities with the binding pocket residues of HIV-1 PR. The goal for this inhibitor was that the Hyp residue will push the coordinated water molecule out of the binding pocket thereby eliminating the threat of hydrolysis of the substrate.

Along the same vein, another peptide inhibitor utilizing the sequence Hyp-Hyp-Lys was considered (Figure 3.8b). The same principles as discussed for the Pro-Hyp-Lys sequence apply. Our supposition was that the Phe-Hyp will also be recognized by HIV-1 PR as it is very similar structurally to Phe-Pro. The addition of the second OH group

present on Hyp allows another opportunity for H-bonding and given that we do not know exactly where either hydroxyl group will actually reside the switch from Pro to Hyp is simply a way of determining if one of those positions will have more favorable interactions than the other. Figure 3.8 shows both transition state isosteres.

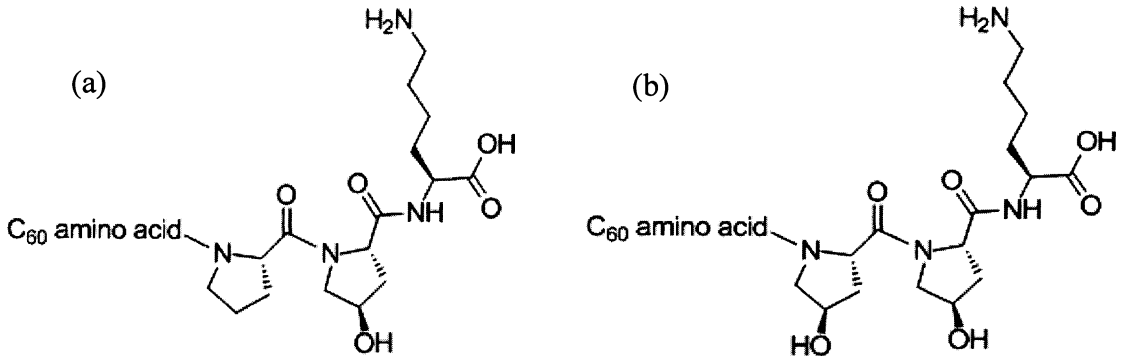


Figure 3.8. The structures of the peptide tails of our transition state isosteres (a) Pro-Hyp-Lys and (b) Hyp_2-Lys.

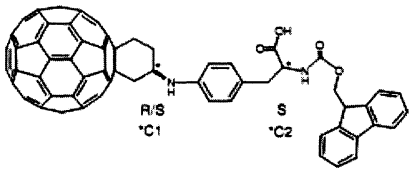
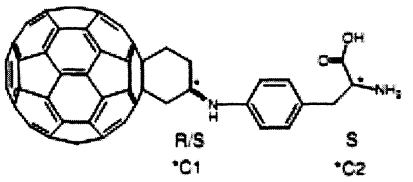
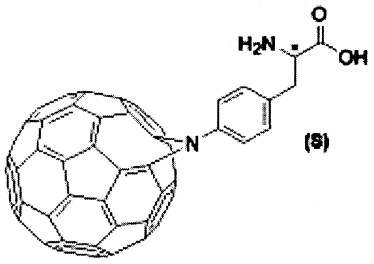
Results and Discussion

The use of fullerene amino acids and peptides as inhibitors of HIV-1 PR was investigated through MD simulations and biological assays. The computational and biological methods utilized by our collaborators are reported elsewhere.[25] Briefly, the MD simulations were performed using a crystal structure of haloperidol, a small molecule inhibitor of HIV-1 PR, in the binding pocket of HIV-1 PR (PDB code:1AID). All of the waters of crystallization remained throughout the simulation but are freely rotatable and replaceable by the ligand, i.e., the inhibitor.

In our preliminary examination of fullerene amino acids as inhibitors of HIV-1 PR we investigated a number of factors, including amino acid core and protecting groups, to determine if subtle changes in the C_{60} amino acid base would have an effect on the observed inhibition. Several of the compounds assayed are illustrated in Table 3.2.

Many of the compounds tested are not the subject of this thesis but are reported in Appendix A to illustrate the quality of the data obtained by this computational method. The synthesis of the compounds was reported elsewhere.[27] Experimental binding affinities (K_i) are converted into binding energy for comparison to simulated values by the equation $\Delta G = - RT \ln K_i$.

Table 3.2. Selected inhibition and MD simulation data from collaborators.

Structure and name	Experimental binding affinity, K_i (nM)	Experimental binding energies (kJ/mol)	Simulated binding energies glide/IFD (kJ/mol)
 Fmoc-Baa	36	-42.75	-46.69
 Baa	124	-39.67	-38.28
 Phe(4-aza-C_{60})-OH (**2.13**)	119	-39.77	-40.00

A plot of the simulated binding energies versus the experimental binding energies showed moderately good agreement between the predicted value and the experimental value ($R^2 = 0.82$, Figure 3.9). While the modeling program could not be considered to be predictive based on the agreement between experimental and simulated values, the data provided by the MD simulations is a good starting point in determining whether or not the compounds are worthwhile to synthesize.

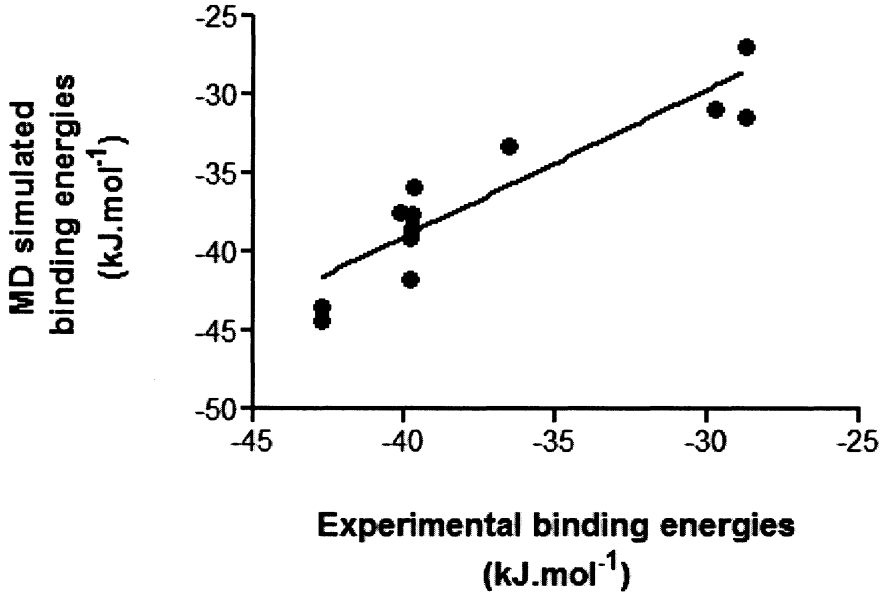


Figure 3.9. The linear regression analysis illustrating the agreement between simulated binding energies and experimental binding energies ($R^2 = 0.82$).

Based on the results, we were able to come to some conclusions about our fullerene amino acid inhibitors. The most important observation during these preliminary experiments was that the presence of the Fmoc protecting group on Baa improved the inhibition dramatically as compared to the corresponding free amine Baa. Initially this was surprising but an inspection of the binding pose of Fmoc-Baa with HIV-1 PR showed that the large aromatic Fmoc group makes favorable interactions with the hydrophobic

binding pocket (Figure 3.10). Also determined from these experiments was the fact that the Diels-Alder fullerene amino acid, Baa, and the newly reported Phe(4-aza/aziridino-C_{60})-OH, **2.13**, had nearly equivalent binding affinities. Given that Baa and **2.13** showed similar inhibition, we speculated that the corresponding Fmoc derivatives (Fmoc-Baa and **2.6**) would follow the same trend. Therefore, we used Fmoc-Phe(4-aza/aziridino-C_{60})-OH (**2.6**) as the fullerene amino acid base for our peptide inhibitors due to its ease of preparation.

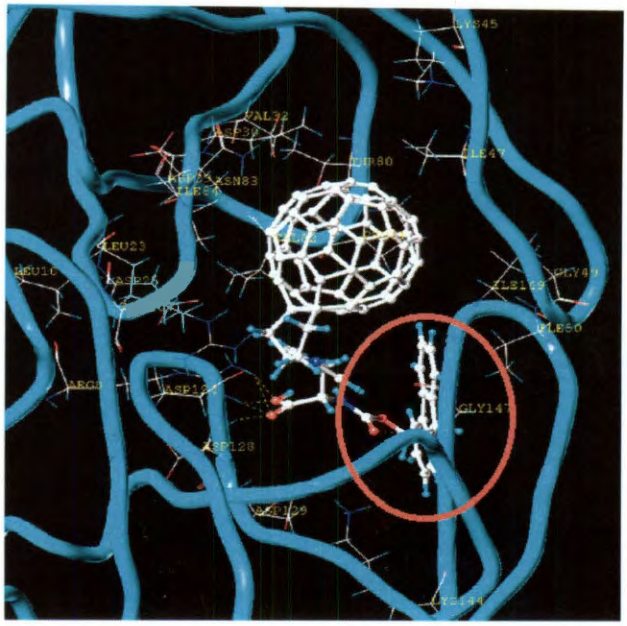


Figure 3.10. Binding pose of Fmoc-Baa with HIV-1 PR. The red circle highlights the Fmoc protecting group.[25]

All C_{60} peptide inhibitors were made through solid phase peptide synthesis (SPPS). The peptide tail was purchased on the resin in its crude form. Compound **2.6** was coupled to the peptide in the dark through standard Fmoc procedures. Due to the acid sensitivity of **2.6**, a solution of 20:2.5:77.5 TFA:triisopropylsilane (TIPS):CH_2Cl_2 was used to cleave the completed peptide sequence off the resin rather than the customary 95:2.5:2.5 TFA:TIPS:H_2O. Figure 3.11 shows the final structure of the three peptide

inhibitors and their MD simulated K_i. Figure 3.12 shows the binding pose of compounds **3.1** with HIV-1 PR as predicted by the MD simulation.

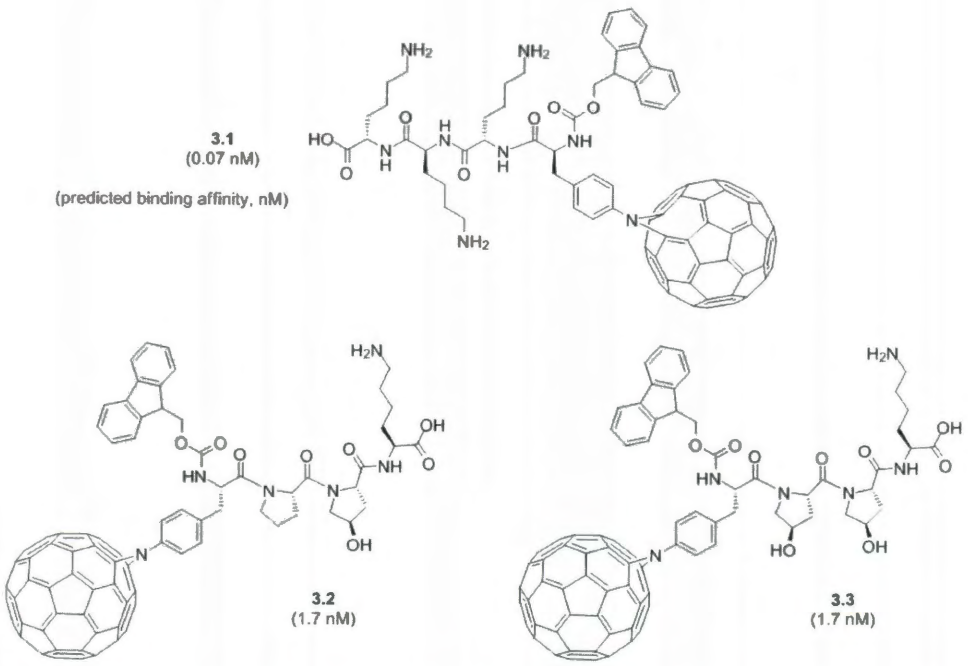


Figure 3.11. The structures of the C_{60}-peptides designed as inhibitors of HIV-1 PR are shown above. In parentheses is the binding affinity as predicted by MD simulations.

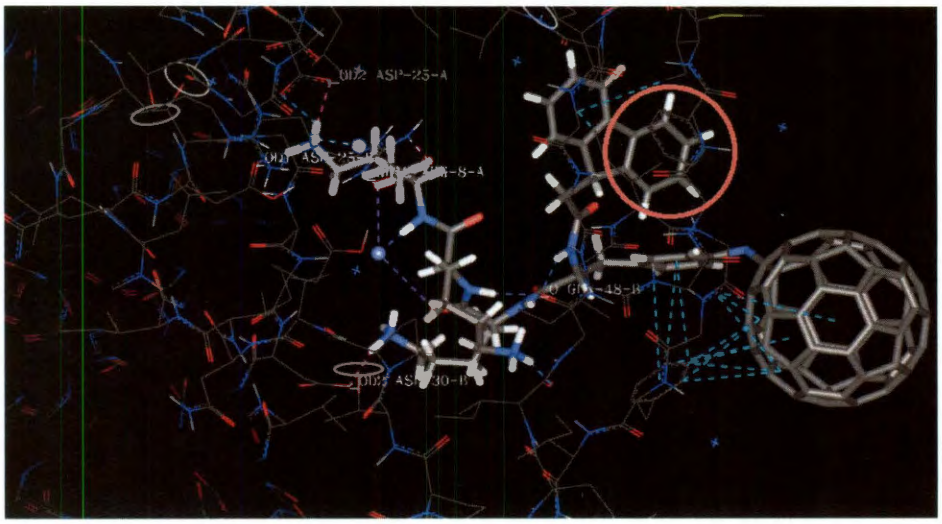


Figure 3.12. Binding pose of compound **3.1** as predicted by MD simulations. The red circle highlights the interaction of the Fmoc protecting group with a Phe residue.

The acidic cleavage from the SPPS resin results in a small degree of C_{60} loss and some conversion of the aza linkage to a primary amine. These were the minor impurities present by ESI/ToF MS in the peptide inhibitors post cleavage. The purity of the peptides cleaved from the resin was determined by HPLC with detection at 220 for the peptide and 330 nm for C_{60}. Figure 3.13 shows an HPLC chromatogram of **3.1** illustrating their purity after cleavage from the resin. Based on ESI MS, the low retention time impurities are the amino peptide minus C_{60} and uncoupled peptide.

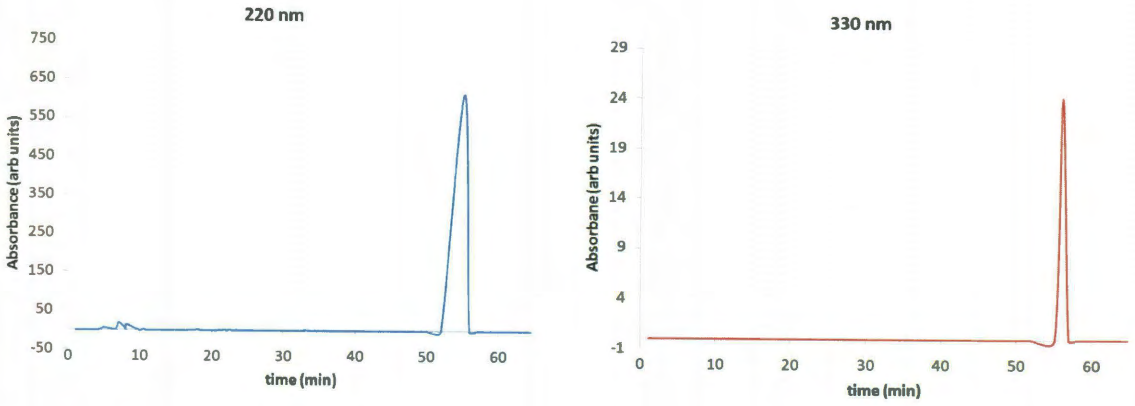


Figure 3.13. Typical HPLC chromatograms of the C_{60} peptides, with detection at 220 nm for the peptide moieties and 330 nm for the C_{60} moiety, illustrating the purity of the peptides after cleavage from the resin.

Due to the tendency of the C_{60} peptides to stick to any sort of column packing material, column purification was avoided when possible. Instead, peptide inhibitors were dissolved in DMSO at a concentration of ~500 μM and filtered through a C_{18} coated SiO_2 plug to remove any C_{60} present. The only impurity remaining was the amino peptide minus C_{60}, which has absorbance at 220 nm (peptide bond) and 256 nm (Phe residue). Therefore, the concentration of the C_{60} peptide inhibitors used in the assays was determined by UV at 330 nm (ε 0.0383 $\mu M^{-1}cm^{-1}$) to eliminate any absorption from amino peptide minus C_{60} (absorbance 256 nm).

To determine the inhibition constant of our fullerene PRIs we utilized a FRET based assay to evaluate the enzyme kinetics of HIV-1 PR at various concentrations of inhibitors. To evaluate the K_m of our FRET substrate with HIV-1 PR we monitored the substrate hydrolysis at various substrate concentrations in the absence of inhibitors. The $[E_t]$ was determined by titration with a tight-binding inhibitor. The K_m of our substrate was determined to be approximately 1 μM and the $[E_t]$ used in the assays was determined to be 37.5 nM (Experimental). In addition to the kinetic constant K_m and the $[E_t]$, a series of control experiments were also performed (Table 3.3).

Table 3.3. Control experiments performed before PRI analysis.

Control	Description	Determination
positive	no test compound	uninhibited reaction velocity, v_o
inhibitor	pepstatin A	known inhibitor of HIV-1 PR
vehicle	DMSO	inhibition due to vehicle
test compound	assay buffer, no enzyme	autofluorescence of C_{60} PRIs
substrate	assay buffer, no enzyme	baseline
C_{60} amino acid	Fmoc-Phe(4-NH_2)-OH	inhibition due to C_{60}
C_{60} peptide	Fmoc-Phe(4-NH_2)-Pro-Hyp-Lys	inhibition due to C_{60}
hydrophobic	Stearic acid-Lys_3	inhibition due to hydrophobicity

The inhibition constants of several amino acid PRIs as determined by cell-free FRET based assay are given in Table 3.4 along with their simulated binding scores for comparison. The K_i for pepstatin A by this analysis method was 420 nM (literature value 500 nM).[28] All of the C_{60} containing samples showed better inhibition than the non-C_{60} control by ~1000 fold. The C_{60} control, Fmoc-Phe(4-NH_2)-OH, had in inhibition constant of >10 μM, in essence, no inhibition in comparison with any of the fullerene amino acids,

which all exhibited at least nM inhibition. Inhibition constants of >10 μM are outside the determinable range by this analysis method. From this we can conclude that it is the presence of C_{60} that is responsible for the observed inhibition. Both of the Fmoc protected amino acids (Fmoc-Baa, 36 nM, and **2.6**, 110 nM) showed improved inhibition over their free amine counterparts (Baa, 119 nM, and **2.14**, 120 nM) supporting our hypothesis that the Fmoc protecting group forms favorable interactions with the hydrophobic binding pocket. In fact, any protecting group on the amino moiety improved the observed inhibition in comparison to the free amine, as evidenced by the Boc-Baa derivative, K_i 80 nM. However, compound **2.6**, Fmoc-Phe(4-aza-C_{60})-OH, was about three times less potent than the analogous Fmoc-Baa despite the fact that their deprotected analogs showed nearly equivalent inhibition.

Table 3.4. Inhibition constants of C_{60} amino acid PRIs and controls.

Compound	Inhibition constant, K_i (nM)	Experimental binding energy ($kJmol^{-1}$)	Simulated binding energy ($kJmol^{-1}$)
Fmoc-Phe-OH	> 10000	-24.9	-
pepstatin A	420	-36.3	-
Boc-Baa	80.	-40.5	-
Fmoc-Baa[a]	36	-42.8	-46.69
Baa[a]	119	-39.7	-38.83
2.6	110	-39.7	-49.47
2.14[a]	120	-39.8	-41.75

[a] Indicates collaborators results shown for comparison to other fullerene inhibitors.

The inhibition constant of the C_{60} peptides and controls as determined by FRET based assay are given in Table 3.5 in addition to the simulated binding energies for comparison. The most notable observation from the peptide experiments was that the predictive modeling program highly overvalued the inhibition of all of the peptide inhibitors while the more simple amino acid PRIs were well estimated. A possible source of the discrepancy is that the peptide inhibitors are much more flexible and, particularly in the case of the Lys residues, have many possible conformations of the R sidechains. In addition to the flexibility, the peptide inhibitors are much larger and possibly not well represented by the haloperidol crystal structure used in the MD simulations. In this regard, a crystal structure of HIV-1 PR with a peptide inhibitor, such as pepstatin A, may provide a more accurate starting point for the MD simulations of the larger peptide inhibitors.

Table 3.5. Inhibition constants of C_{60} peptides and control.

Compound	Inhibition constant, K_i (nM)	Experimental binding energy ($kJmol^{-1}$)	Simulated binding energy ($kJmol^{-1}$)
pepstatin A	420	-36.3	-
2.6	110	-39.7	-49.47
3.1[a]	85	-40.3	-58.32
3.2	120[†]	-39.4	-50.38
3.3	76	-40.6	-50.30
SA-Lys_3-OH (**3.4**)[a]	6310	-29.7	-
Fmoc-Phe-Pro-Hyp-Lys (**3.5**)	> 10000	-23.8	-

[a] Indicates collaborators results shown for comparison to other fullerene inhibitors.

[†] Indicates result from single data set.

In our investigation into the inhibition of HIV-1 PR with fullerene peptides we wanted to determine if just the presence of a hydrophobic moiety on a peptide sequence would have an effect on the observed inhibition of PR. In this regard, there was no suitable molecule that can mimic the hydrophobicity, size, and shape of C_{60}. Stearic acid (SA) has been shown to be membranotropic in much the same way as C_{60},[29] and for that reason, it was chosen as a hydrophobic control. The SA-Lys_3 control (**3.4**) showed inhibition of approximately 6.3 μM, better than other controls but still well below the potency of the C_{60} peptides. The inhibition constant of another non-C_{60} control peptide, Fmoc-Phe(4-NH_2)-Pro-Hyp-Lys (**3.5**), was determined to be >10 μM, demonstrating again that the observed inhibition of the reaction velocity was due to C_{60} and not a result of the dipeptide recognition element, Phe-Pro.

All of the peptide inhibitors showed improved inhibition over their C_{60} amino acid base; however, none were as effective as our most potent inhibitor, Fmoc-Baa (K_i 36 nM). The most active of the peptide inhibitors was compound **3.3** with two hydroxyproline residues, K_i 76; compound **3.3** was much more active than compound **3.2**, K_i 120, with only one hydroxyproline residue. It is unknown if the observed increase in binding affinity is due to better placement of the second hydroxyl moiety in regards to the aspartyl residues or if it is due to the increase in H-bonding opportunities; however the ΔΔG is worth approximately that of a H-bond. Regardless, the Phe-Hyp dipeptide sequence does seem to be recognized by HIV-1 PR as we hypothesized. Compound **3.1**, the poly-Lys peptide, also showed improved inhibition over the C_{60} amino acid base. However, without a crystal structure of compound **3.1** in the binding pocket it is impossible to determine if the positively charged Lys residues are forming a salt bridge or if the improved affinity is due to H-bonding. Calculation of the ΔΔG between **3.1** and **2.6** reveals that a salt bridge, worth about 12-17 $kJ.mol^{-1}$, is probably not the cause of the increase, which was approximately – 0.6 $kJ.mol^{-1}$.

In comparison to currently marketed protease inhibitors, none of ours have yet achieved sub-nM levels of inhibition and therefore are not yet viable drug candidates. However, our results with both the amino acids and the peptides are encouraging indicators that the achieved inhibition may be tuned with minor alterations to both the peptide and amino acid structure. It should also be noted that almost all of our fullerene inhibitors show higher levels of inhibition in comparison with the most potent previously reported fullerene derivative (103 nM).

Given that all of the peptide sequences improved the inhibition of the base fullerene amino acid, it is reasonable to assume that they will do the same for our other C_{60} amino acids. Since our most potent inhibitor was Fmoc-Baa, in future studies it should be used as the C_{60} amino acid base despite the length of time required to make it. In addition to the change in the amino acid base, in future studies the length of the peptides should be increased by one Lys residue or more. Initially, when designing the peptides we used Lipinsky's rule of five for drug compounds as a guideline. All of our designed peptide inhibitors violated these rules at least partially and therefore a minimum sequence length was used for them all. However, HIV-1 PR recognizes peptides of at least five residues most efficiently.[15] We know that cationic fullerene peptide sequences cross a cellular membrane easily,[30] therefore we surmise that we would not sacrifice membrane permeability by increasing the number of Lys residues on the tail and could potentially improve the molecular recognition with HIV-1 PR with the addition. Future work should maintain the Hyp_2 dipeptide subunit but should also include two or more Lys residue on the C-terminal end. This structural modification should improve solubility, make the peptide more recognizable by PR, and improve the affinity through H-bonding.

The unattainable goal thus far in the field of HIV-1 PR inhibition is an inhibitor that binds equally well to the mutated strains of PR as to the native form. Given that point mutations usually occur within the hydrophobic binding pocket, and those mutations are

still hydrophobic residues, it is reasonable to infer that the non-specific hydrophobic interaction of C_{60} with the binding pocket of mutated strains will still be strong. Other questions about the specific interactions of C_{60} with HIV-1 PR could be determined by a crystal structure of one of the C_{60} inhibitors, such as Fmoc-Baa, in the binding pocket.

Conclusion

We have reported the synthesis of three, novel fullerene peptide sequences and investigated their inhibition of HIV-1 PR. We have determined that C_{60} amino acids are a good base structure in the design of protease inhibitors and that their inhibition can be improved upon by the addition of designer peptide sequences. We found that the most potent of the C_{60} based inhibitors was Fmoc-Baa (K_i 36 nM) due to the electrostatic interactions of both the Fmoc protecting group and C_{60}. Of the peptide inhibitors, the Hyp-Hyp-Lys (K_i 75.1) sequence showed the highest level of inhibition and improved the inhibition of its C_{60} amino acid base, **2.6**, (K_i 108). Based on these findings, we can reason that HIV-1 PR may be able to recognize the Phe-Hyp dipeptide sequence; however, it is not known if that recognition would result in specificity to HIV-1 PR over cellular proteases. Given the improvement in the observed protease inhibition achieved by the addition of the peptide tail to compound **2.6**, it is reasonable to assume that the inhibition of Fmoc-Baa may also be improved by such an addition. While our C_{60} peptide inhibitors are not yet viable drug candidates, the results of these experiments are promising.

Experimental

The FRET peptide substrate (sequence not given), HIV-1 PR, and assay buffer A (Tris, Glycerol, KCl, DTT, and EDTA with 500mM KCl, pH 4.7) were provided in the ProAssay^{TM} HIV- Protease kit from ProteinOne. The peptide sequence of the FRET

peptide substrate is derived from the native p17/p24 cleavage site on Prgag (native substrate) for HIV-1 protease. Protected amino acids were purchased from Chem-Impex International, Inc., unless otherwise specified, and used without further purification. All inhibitor peptide sequences were purchased from BioSynthesis, Inc. in the crude form on Wang resin. Solvents were purchased from Sigma Aldrich and used without further purification. All other chemicals and reagents were purchased from Sigma-Aldrich. High performance liquid chromatography (HPLC) was performed on a Varian ProStar instrument using a Varian Dynamax C_{18} column (250 x 21.6 mm) for purification when necessary. Mass spectrometry measurements were performed on a Bruker AutoFlex II/ToF/ToF and a Bruker Electrospray Ionization/μToF. α-Cyano-4-hydroxycinnamic acid were used as the matrix to characterize all peptides by MALDI/TOF (positive mode). UV-vis spectra were recorded on a Varian Cary 5000 spectrometer. Fluorescence data was recorded on a BioTek Synergy 4 fluorescence platereader.

General fullerene peptide synthesis. All peptides were purchased on the resin with orthogonal protection for Fmoc SPPS on the side chains and N-terminus. The fullerene amino acid, or a control amino acid, was coupled to the peptide sequence through published literature methods in a 3:1 amino acid to peptide ratio.[31] The reaction mixture was shielded from light throughout the synthesis and purification. Peptides were cleaved from the resin using 20% TFA with 2.5% TIPS in DCM and precipitated with cold diethylether then washed with ether (3 x 50 mL). The crude peptide was dissolved in an appropriate solvent or buffer and mixed with a small amount of C_{18} coated silica gel and immediately filtered to remove the salts for analysis then lyophilized and stored under vacuum in the dark. When necessary the peptide was purified by HPLC (C_{18} prep scale, 95:5 to 5:95 0.1% formic acid and isopropanol, 65 min). The major impurity present is C_{60} by ESI MS.

Fmoc-Phe(4-aza-C_{60})-Lys_3-OH (3.1). The crude peptide, Fmoc-Lys(Boc)-[Lys(Boc)]$_2$-Wang resin (215.5 mg, 0.2 mmol/g_{resin}), was added to a 10 mL SPPS synthesis vessel. DMF was used to swell the resin for 1 h then filtered. The Fmoc protecting group on the terminal Lys residue was removed with 25% piperidine in DMF (2 x 10 mL) afterwhich the resin was rinsed with DMF (3 x 10 mL). Compound **2.5a** (150 mg, 0.132 mmol), HOBt (17.81 mg, 0.132 mmol), PyBop (68.64 mg, 0.132 mmol), and diisopropylethylamine (DIEA, 68.9 μL, 0.396 mmol) were dissolved in DMF:CH_2Cl_2 (1:2, 8 mL) and sonicated for two minutes then added to the resin. The synthesis vessel was always kept in the dark and was mixed for 24 h. The excess reagents were filtered away and the resin was washed with CH_2Cl_2 (3 x 10 mL), and DMF (3 x 10 mL) then the resin was shrunk with MeOH (3 x 10 mL). The peptide was cleaved from the resin with 20:2.5:77.5 TFA:TIPS:CH_2Cl_2 and precipitated and washed with the addition of cold Et_2O (3 x 50 mL). ESI/ToF shows the major impurity is C_{60}. To remove C_{60} the peptide is dissolved in DMSO and filtered through a C_{18} silica plug. The product was lyophilized to dryness. Yield: 33 mg, 50%. ESI/Tof (M+1) requires *m/z* 1505., found *m/z* 1505.45.

Fmoc-Phe(4-aza-C_{60})-Pro-Hyp-Lys (3.2). The crude peptide, Fmoc-Pro-Hyp(tBu)-Lys(Boc)-Wang resin (102.63, 0.217 mmol/g_{resin}, 0.022 mmol), was added to a 10 mL SPPS synthesis vessel. DMF was used to swell the resin for 1 h then filtered. The Fmoc protecting group on the terminal Lys residue was removed with 25% piperidine in DMF (2 x 10 mL) afterwhich the resin was rinsed with DMF (3 x 10 mL). Compound **2.5a** (75 mg, 0.067 mmol), HOBt (9.04 mg, 0.067 mmol), PyBop (34.83 mg, 0.067 mmol), and diisopropylethylamine (DIEA, 35 μL, 0.201mmol) were dissolved in DMF:CH_2Cl_2 (1:2, 8 mL) and sonicated for two minutes then added to the resin. The synthesis vessel was always kept in the dark and was mixed for 24 h. The excess reagents were filtered away and the resin was washed with CH_2Cl_2 (3 x 10 mL), and DMF (3 x 10 mL) then the resin was shrunk with MeOH (3 x 10 mL). The peptide was cleaved from

the resin with 20:2.5:77.5 TFA:TIPS:CH_2Cl_2 and precipitated and washed with the addition of cold Et_2O (3 x 50 mL). The peptide was pure as cleaved off the resin by ESI/ToF with a minor C_{60} impurity. The product was dissolved in MeCN:H_2O (1:1, 1.0 mL) and lyophilized to dryness. Yield: 53 mg, 80%. ESI/Tof (M+1) requires *m/z* 1459.41, found *m/z* 1459.34.

Fmoc-Phe(4-aza-C_{60})-Hyp-Hyp-Lys (3.3). Prepared in a similar manner to compound **3.2** using the crude peptide, Fmoc-Hyp(tBu)-Hyp(tBu)-Lys(Boc)-Wang resin (100.40 mg, 0.22 mmol/g_{resin}, 0.022 mmol), compound **2.5a** (75 mg, 0.067 mmol), HOBt (9.04 mg, 0.067 mmol), PyBop (34.83 mg, 0.067 mmol), and diisopropylethylamine (DIEA, 35 μL, 0.201mmol). Yield: 27.6 mg, 85%. ESI/Tof (M+1) requires *m/z* 1475.9, found *m/z* 1476.23.

Stearic Acid-Lys_3 (3.4). Prepared in a similar manner to compound **3.1** using the crude peptide, Fmoc-Lys(Boc)-[Lys(Boc)]$_2$-Wang resin (74.88 mg, 0.266 mmol/g_{resin}, 0.02 mmol), stearic acid (17 mg, 0.06 mmol), HOBt (8.07 mg, 0.06 mmol), PyBop (31.10 mg, 0.06 mmol), and diisopropylethylamine (DIEA, 31.2 μL, 0.179 mmol). Yield: 12.05 mg, 90%. ESI/Tof (M+1) requires *m/z* 669.56, found *m/z* 669.54.

Fmoc-Phe(4-NH_2)-Pro-Hyp-Lys (3.5). Prepared in a similar manner to compound **3.2** using the crude peptide, Fmoc-Pro-Hyp(tBu)-Lys(Boc)-Wang resin (71.6 mg, 0.22 mmol/g_{resin}), Fmoc-Phe(4-NH_2)-OH (20.0 mg, 0.047 mmol), HOBt (6.3 mg, 0.047 mmol), PyBop (24.3 mg, 0.047 mmol), and diisopropylethylamine (DIEA, 24.4 μL, 0.140 mmol) Yield: 11 mg, 90%. ESI/Tof (M+1) requires *m/z* 767.91, found *m/z* 767.34.

Determination of substrate K_m. The K_m of a substrate is determined by evaluating the initial velocity of substrate hydrolysis at various substrate concentrations in the absence of an inhibitor. The concentrations of substrate assayed by this method were from 16.5 – 0.2 μM. Substrate hydrolysis was followed as an increase in fluorescence (excitation/emission 490/530 nm). Data was collected at 25 °C for one hour, or as long as the data remained linear, at one minute increments. The total reaction volume was 25 μL. The kinetic parameters K_m and V_{max} were determined by fitting the reaction progress curve to the M-M equation. Figure 3.10 shows the Michaelis-Menten plot of our enzyme kinetic data with HIV-1 PR (red curve). All data presented was the average of at least duplicate measurements. The K_m of our substrate was determined to be 0.6 μM. All future kinetic inhibition assays were performed at a substrate concentration of ~1 μM. We confirmed that our enzyme was following Michaelis-Menten kinetics by fitting the data to the equation for a standard rectangular hyperbola (Figure 3.14, green curve).

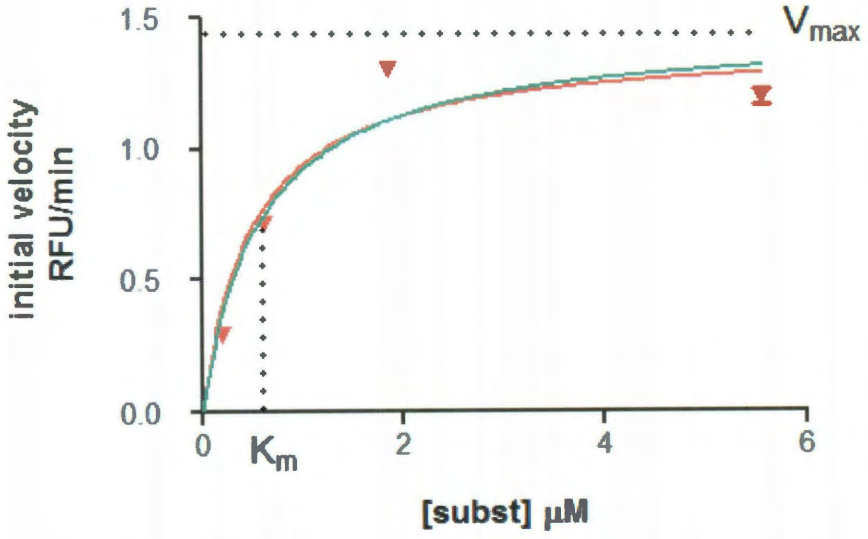


Figure 3.14. The red curve is the plot of M-M type kinetics observed during the hydrolysis of the FRET substrate by HIV-1 PR where V_{max} = 1.459 and K_m = 0.58 μM. The green curve represents the same data fit to the equation of a standard hyperbola.

Determination of total active site concentration [E_t]. The total active site concentration values were determined by a cell-free FRET based assay. Typically, 1 pmol (12 µL, 0.09 µM) of HIV-1 PR was added to 13 µL of dilution buffer A containing the substrate at a concentration near its K_m (1 µM) and various concentrations of a tight binding inhibitor dissolved in DMSO (final concentrations from 100 µM to 0.01 nM). The final concentrations of DMSO were kept below 2.5% (v/v). Substrate hydrolysis was followed as an increase in fluorescence (excitation/emission 490/530 nm). Data was collected at 25 °C for one hour, or as long as the data remained linear, at one minute increments. The total reaction volume was 25 µL. All data presented is the average of at least duplicate measurements. The solubility of the tight binding inhibitor was sufficient for all the concentrations used. The DMSO/water mixture was used in a 1:9 ratio and no precipitation was observed. Linear regression was performed on a plot of fractional velocity (v_i/v_o) versus inhibitor concentration to determine the total active site concentration (Figure 3.15). When the fractional velocity plateaus near zero, the enzyme is said to be completely, or nearly completely, inhibited and the rate of the reaction is determined simply by dissociation of the enzyme-inhibitor complex. The concentration of a tight binding inhibitor is said to be equal to the total active cite concentration, [E_t], at the *x*-intercept of the linear portion of the data, before the fractional velocity plateaus.

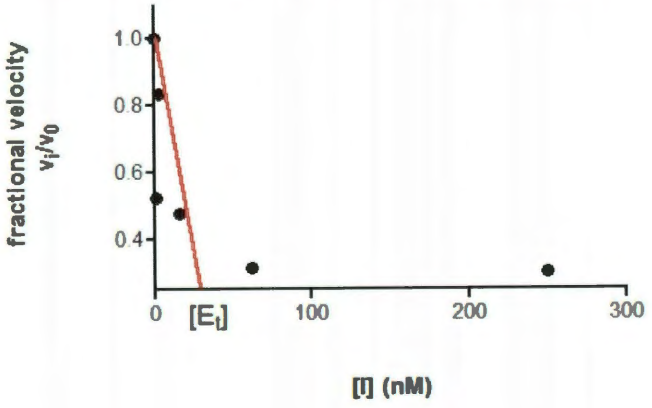


Figure 3.15. Above is the plot of fractional velocity (v_i/v_o) versus inhibitor concentration used to determine the total active site concentration of the enzyme, [E_t] = 37.5 nM.

Inhibition assay. The K_i values were determined by a cell-free FRET based assay using the FRET peptide substrate. Typically, 1 pmol (12 μL, 0.09 μM) of HIV-1 PR was added to 13 μL of dilution buffer A containing the substrate at a concentration near its K_m (1 μM) and various concentrations of the inhibitor dissolved in DMSO (final concentrations from 100 μM to 0.01 nM). The final concentrations of DMSO were kept below 2.5% (v/v). Substrate hydrolysis was followed as an increase in fluorescence (excitation/emission 490/530 nm). Data was collected at 25 °C for one hour, or as long as the data remained linear, at one minute increments. The total reaction volume was 25 μL. The solubility of the compounds was sufficient for all the concentrations used. The DMSO/water mixture was used in a 1:9 ratio and no precipitation was observed. Inhibition data was analyzed using the equation for competitive inhibition according to Williams and Morrison.[32] The data was fit using non-linear regression with the Graphpad Prism software.[33]

References

1 M. Prato, A. Bianco, M. Maggini, G. Scorrano, C. Toniolo, and F. Wudl, *J. Org. Chem.*, 1993, **58**, 5578.

2 J. Yang, L. B. Alemany, J. Driver, J. D. Hartgerink, and A. R. Barron, *Chem. Eur. J.*, 2007, **13**, 2530.

3 (a) M. E. Vol'pin, Z. N. Parnes, and V. S. Romanova, *Russ. Chem. Bull.*, 1998, **47**, 1021. (b) J. Yang, K. Wang, J. Driver, J. Yang, and A. R. Barron, *Org. Biomol. Chem.*, 2007, **5**, 260. (c) I. Andreeva, A. Petrukhina, A. Garmanova, A. Babakhin, S. Andreev, V. Romanova, P. Troshin, O. Troshina, and L. DuBuske, *Fullerenes, Nanotubes, Carbon Nanostruct.*, 2008, **16**, 89. (d) F. Pellarini, D. Pantarotto, T. Da Ros, A. Giangaspero, A. Tossi, and M. Prato, *Org. Lett.*, 2001,

3, 1845. (e) A. Bianco, T. Da Ros, M. Prato, and C. Toniolo, *J. Peptide Sci.*, 2001, **7**, 208. (f) D. Pantarotto, N. Tagmatarchis, A. Bianco, and M. Prato, *Mini-Rev. Med. Chem.*, 2004, **4**, 805.

4 A. Brik and C.-H. Wong, *Org. Biomol. Chem.*, 2003, **1**, 5.

5 (a) R. A. Kramer, M. D. Schaber, A. M. Skalka, K. Ganguly, F. Wong-Staal, and E. P. Reddy, *Science*, 1986, **231**, 1580. (b) J. R. Huff, *J. Med. Chem.*, 1991, **34**, 2305.

6 D. R. Davies, *Annu. Rev. Biophys. Chem.*, 1990, **19**, 189.

7 M. A. Navia, P. M. D. Fitzgerald, B. M. McKeever, C.-T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke, and J. P. Springer, *Nature*, 1989, **337**, 615.

8 J. D. Bernal, and D. Crowfoot, *Nature*, 1934, **133**, 794.

9 M. N. G. James (ed.), *Aspartic Proteinases: Retroviral and Cellular Enzymes*, Plenum Press, New York, 1998, 9.

10 R. E. Babine and S. L. Bender, *Chem. Rev.*, 1997, **97**, 1359.

11 L. J. Hyland, T. A. Tomaszek, and T. D. Meek, *Biochemistry*, 1991, **30**, 8454.

12 http://www.doe-mbi.ucla.edu/~yeates/153AH_2009_project/sriphanlop.html. Accessed 8/9/2010.

13 (a) K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson, and D. R. Davies, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6612. (b) M. Jaskólski, A. G. Tomasselli, T. K. Sawyer, D. G. Staples, R. L. Heinrikson, J. Schneider, S. B. H. Kent, and A. Wlodawer, *Biochemistry*, 1991, **30**, 1600.

14 A. Wlodawer and J. Vondrasek, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 249.

[15] R. A. Copeland, *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*, Wiley, 2005, Hoboken, New Jersey, 178.

[16] (a) www.fda.gov. Accesssed 8/15/2010. (b) O. Aruksakunwong, S. Promsri, K. Wittayanarakul, P. Nimmanpipug, V. S. Lee, A. Wijitkosoom, P. Sompornpisut, and S. Hannongbua, *Curr. Comput.-Aided Drug Des.*, 2007, **3**, 201. (c) E. Lefebvre and C. A. Schiffer, *AIDS Rev.*, 2008, **10**, 131.

[17] S. H. Friedman, D. L. DeCamp, R. P. Sijbesma, G. Srdanov, F. Wudl, and G. L. Kenyon, *J. Am. Chem. Soc.*, 1993, **115**, 6506.

[18] (a) S. Bosi, T. Da Ros, G. Spalluto, J. Balzarini, and M. Prato, *Bioorg. & Med. Chem.*, 2003, 4437. (b) D. I. Schuster, L. J. Wilson, A. N. Kirschner, R. F. Schinazi, S. Schlueter-Wirtz, P. Tharnish, T. Barnett, J. Ermolieff, J. Tang, M. Brettreich, and A. Hirsch, A. In *Fullerene 2000—Functionalized Fullerenes*; N. Martin, M. Maggini, D. M. Guldi, Eds., The Electrochemical Society: Pennington, NJ, USA, 2000; 9, 267. (c) S. Marchesan, T. Da Ros, G. Spalluto, J. Balzarini, and M. Prato, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3615.

[19] S. H. Friedman, P. S. Ganapathi, Y. Rubin, and G. L. Kenyon, *J. Med. Chem.*, 1998, **41**, 2424.

[20] A. G. Marangoni, *Enzyme Kinetics: A Modern Approach*, Wiley, 2003, Hoboken, New Jersey, 44.

[21] S. G. Waley, *Biochem J.*, 1993, **294**, 195.

[22] (a) S. M. Rodems, B. D. Hamman, C. Lin, J. Zhao, S. Shah, D. Heidary, L. Makings, J. H. Stack, and B. A. Pollok, *Assay Drug Dev. Technol.*, 2002, **1**, 9. (b) Y.-C. Liu, V. Huang, T.-C. Chao, C.-D. Hsiao, A. Lin, M.-F. Chang, and L.-P. Chow, *Biochem. Biophys. Res. Commun.*, 2005, **333**, 194. (c) G. J. R. Zaman, A.

Garritsen, T. de Boer, and C. A. A. van Boeckel, *Comb. Chem. High Throughput Screening*, 2003, **6**, 313.

[23] User manual for Pro-Assay™ HIV-1 Protease Assay Kit from ProteinOne.

[24] S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, and M. G. Papadopoulos, *J. Chem. Inf. Model.*, 2009, **49**, 1139.

[25] S. Durdagi, T. Mavromoustakos, N. Chronakis, and M. G. Papadopoulos, *Bioorg. Med. Chem.*, 2008, **16**, 9957.

[26] (a) C. A. Lipinski, *Drug Discov Today: Technologies*, 2004, **1**, 337. (b) C. A. Lipinski, *J. Pharmacol. Toxicol. Methods*, 2000, **44**, 235.

[27] (a) J. Yang and A. R. Barron, *Chem. Commun.*, 2004, 2884. (b) J. Z. Yang, Doctoral Thesis, 2005. (c) N. Doostar, Doctoral Thesis, 2010.

[28] E. S. Furfine, *Current Protocols in Pharmacology*, 1998, 3.2.1.

[29] S. Trabulo, A. L. Cardoso, M. Mano, and M. C. Pedroso de Lima, *Pharmaceuticals*, 2010, **3**, 961.

[30] J. Yang, K. Wang, J. Driver, J. Yang, and A. R. Barron, *Org. Biomol. Chem.*, 2007, **5**, 260.

[31] *NovaBiochem Catalog & Peptide Synthesis Handbook*, 1999, NovaBiochem, La Jolla, California.

[32] J. W. Williams and J. F. Morrison, *Methods Enzymol.*, 1979, **63**, 437.

[33] Non-linear regression determination of Morrison K_i performed on GraphPad Prism version 5.03 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

Chapter 4

Synthesis of Highly Functionalized Graphene

Introduction

Recent years have brought about a renewed interest in graphite or specifically individual or few-layer ($n < 5$) sheets of graphite called graphene. The unique physical and electronic properties of graphene make it attractive as a substitute for other more costly carbon nanostructures.[1] Due to its similarities and differences with other carbon nanostructures, graphene (2 dimensional, 2D) is no longer the underappreciated parent of fullerenes (0D) and carbon nanotubes (1D) (Figure 4.1). The ability to individualize sheets of graphene has brought the material to the forefront of carbon research.[2] And while graphene does provide similar scale, properties, and reaction pathways as fullerenes and CNTs, it does so at a fraction of the cost. Regardless of the novel electronic properties of graphene, it also presents an inexpensive, large surface area scaffold for the covalent attachment of organic molecules. Covalent funtionalization provides a robust material for use as a scaffold and as a means of incorporating additional functionality. Incorporation of functionalized graphene sheets into composite materials may enhance mechanical and electronic properties.

The majority of reports of covalently functionalized graphene use graphene oxide (GO) as the starting material.[3] While the structure of GO has been shown to include regions of unoxidized benzene rings, the presence of islands of epoxides and hydroxyl groups on the basal plane and carboxylic acids around the perimeter of the sheet make GO extremely hydrophilic. In addition, graphene oxide contains a large percentage of water requiring more than four weeks to dry post synthesis.[4] While the GO route does result in highly functionalized graphene sheets, the drawbacks are that reactions using GO as the starting material are limited to aqueous or polar conditions and more important a second reduction step is necessary to remove the excess oxides still present.[3] A simpler route to widespread functionalization is therefore desirable. In this regard, we report a new, high yield method of covalently functionalizing graphene sheets.

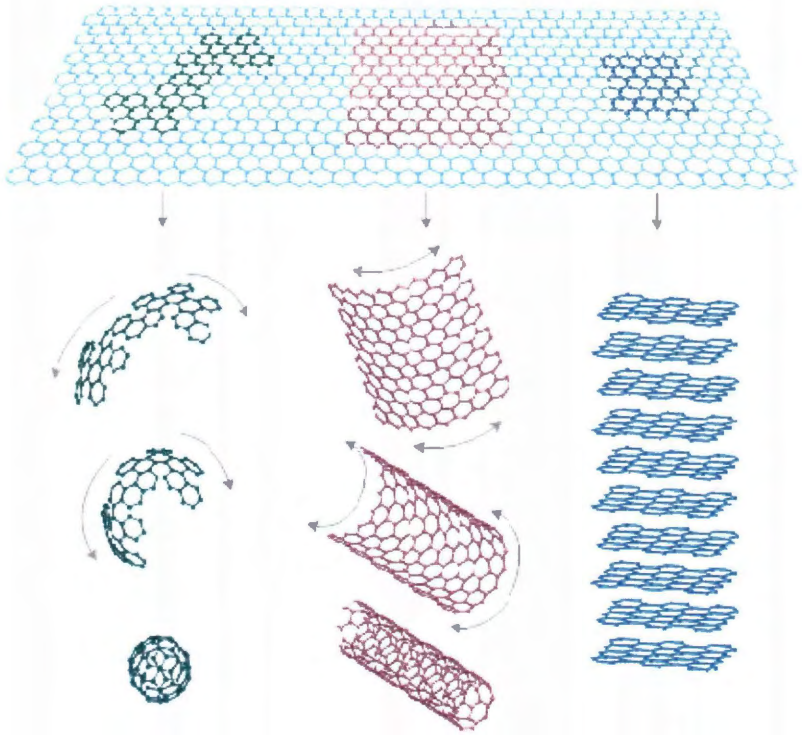


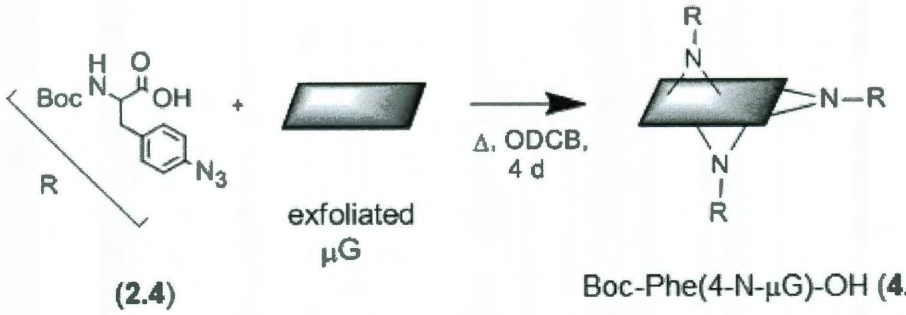
Figure 4.1. Three of the carbon allotropes are presented above.[1]

It has been shown that both CNTs and fullerenes are easily functionalized by nitrene addition, generated *in-situ* by thermal decomposition of azido starting materials.[5] The result is high density functionalization of the nanostructure surface through a [2+1] cycloaddition to the C-C double bonds forming an aziridino-ring linkage, however, this methodology has only been applied to curved or strained (i.e. reactive) carbon nanostructures. The alkyl and aryl azido starting materials are easily synthesized on the gram scale and azides offer a wide range of functional group tolerance.[6] For our purposes, the azido moiety resides on the R side chain of the amino acid phenylalanine (Phe) allowing incorporation of two new chemical handles on the graphene scaffold for subsequent reaction. In this regard, the graphene sheet could be used as a solid support for solid phase peptide synthesis (SPPS) of a sequence on the graphene surface. The addition of sequences of amino acids to the graphene surface should impart water solubility and some level of biocompatibility.

Results and Discussion

The high yield functionalization of graphitic materials through nitrene addition has been investigated. The nitrene was generated by the thermal decomposition of azido starting materials, in this case, a Boc protected amino acid, phenylalanine. The synthetic graphite was exfoliated into o-dichlorobenzene (ODCB) through sonication prior to the nitrene addition.

Microcrystalline graphite was exfoliated in ODCB to make a clear gray solution of graphene sheets without the use of harsh oxidizing conditions.[7] Previous reports using this exfoliation method use high-shear mixing, or homogenization, to further exfoliate the sample before centrifugation. For our future purposes of using the graphene sheet as a support for SPPS, the large lateral dimensions of the starting material were desirable, and therefore, the homogenization step was eliminated. It is worth noting that this reaction was not reproducible using thermally expanded graphite intercalation compounds (GIC)[8] as the starting material. Given that GICs use a variety of reactive molecules to intercalate between graphene layers,[4] side reactions with reactants are inevitable, thus limiting their use as graphitic starting materials as well.



Scheme 4.1. Reaction scheme showing the nitrene addition to exfoliated uG.

In a typical reaction the exfoliated μG solution was stirred with Boc-Phe(4-N_3)-OH (**2.4**) at refluxing temperatures, ~180 °C, for four days. The reaction was filtered over a 0.2 μm PTFE filter paper, and washed to remove any unreacted Phe(4-N_3)-OH (Scheme 4.1). The product **4.1**

was found to make an amber colored solution in DMSO, averaging 0.06 $mg.mL^{-1}$ after 10 min. of probe sonication and centrifugation (3400 RPMs for 15 min). The product also proved sparingly soluble in other organic solvents such as DMF, toluene, and chloroform in comparison to the μG starting material. To confirm the chemical modification FTIR spectroscopy, Raman spectroscopy, and XPS were utilized.

The elemental composition of the product **4.1** was determined by XPS, confirming the presence of the phenylalanine (Figure 4.2). The oxygen (16.1%) and nitrogen (13.5%) content are consistent with the presence of the phenylalanine substituent (Figure 4.3). The XPS data of **4.1** indicates 13:1 carbon to phenylalanine ratio. However, we expected to see approximately a 2:1 oxygen:nitrogen composition instead we observed a 1.2:1 oxygen to nitrogen ratio. We suspected that the discrepancy was due to the loss of the Boc protecting group, which can occur at 180 °C;[9] however, the Fourier transform infrared (FTIR) spectra of **4.1** (Figure 4.4) showed the characteristic absorption of a secondary carbamate carbonyl stretch at 1655 cm^{-1}, indicating there was only partial loss of the Boc protecting group if any. In addition, the spectrum shows absorption in the carboxylic acid (~3200 cm^{-1} stretch and 1438 cm^{-1} bend) and the amide C-N regions (1319 cm^{-1}) relative to the μG starting material. Also present in the spectrum are the NH stretch (3426 cm^{-1}) of the Boc amino protecting group and the prominent C-N amine stretch at 1022 cm^{-1} of the aziridino linkage.

High resolution transmission electron microscopy (TEM) of **4.1** show multi- or few-layer graphene sheets ($n < 5$) layered over each other, with an average length of 300 - 500 nm and with varying widths. The formation of multi- or few- layer sheets is further supported by selected area electron diffraction (SAED, Figure 4.5 inset) and fast Fourier transform (FFT, Figure 4.5) in selected areas. SAED during TEM analysis showed that of a hexagonal close packing (AB) crystallite with six point hexagonal symmetry (Figure 4.5). No ring-like diffraction patterns are observable as would be the case for graphite with random stacking or amorphous carbon; however, additional spots with hexagonal symmetry but misaligned with respect to each other indicated multi-layers.

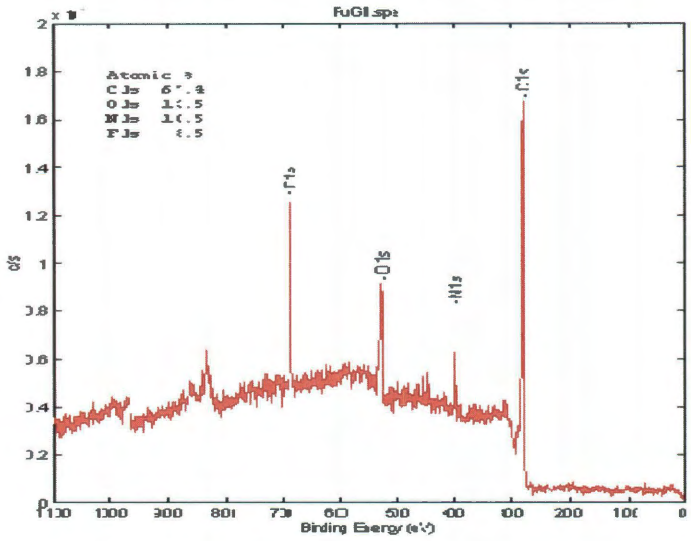


Figure 4.2. Presented above is the XPS survey scan of **4.1** showing the presence of C, O, N, and F (attributed to fluro-grease).

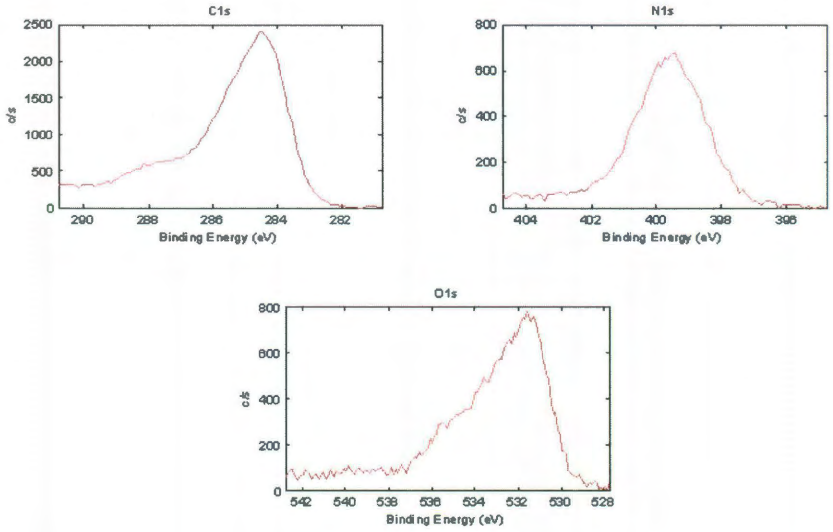


Figure 4.3. The high-resolution XPS C1s, N1s and O1s spectra of **4.1** (y-axis counts/second).

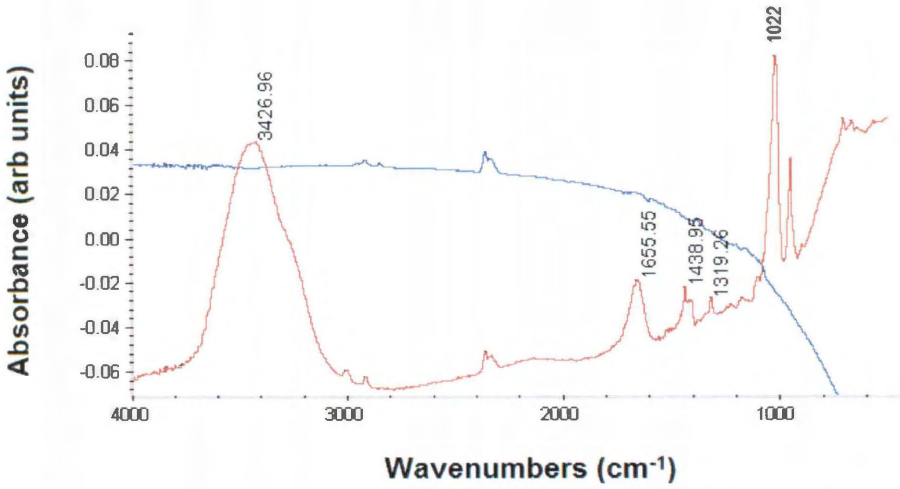


Figure 4.4. Thin film IR spectra of compound **4.1** taken on a KBr salt plate.

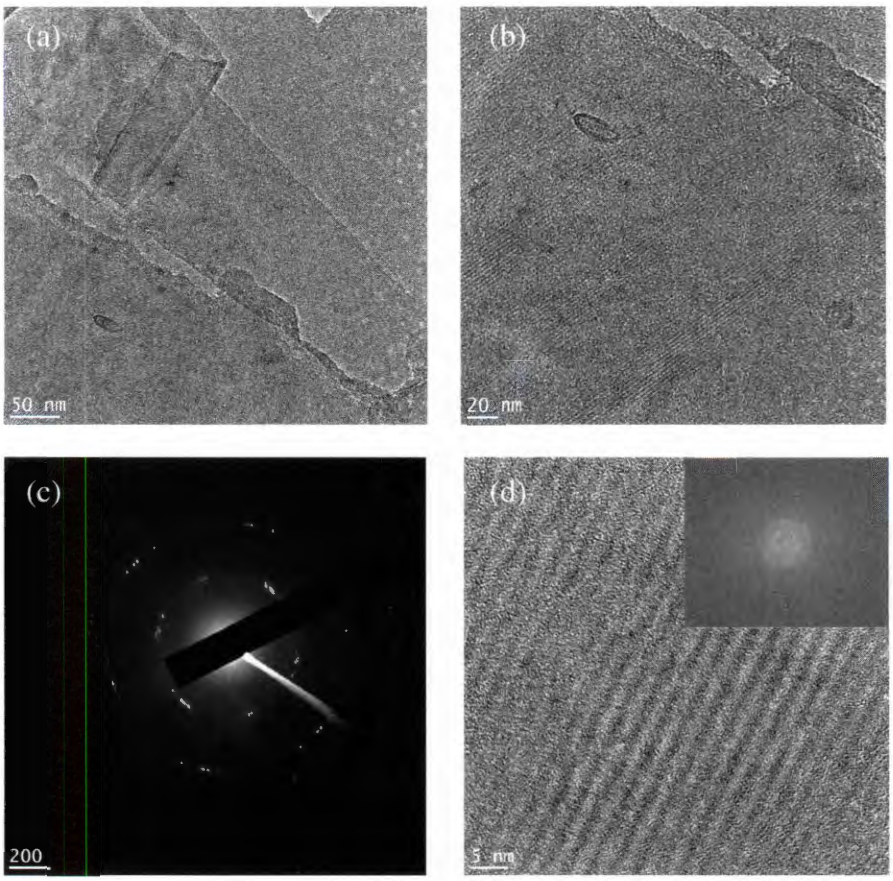


Figure 4.5. TEM micrographs of **4.1** showing (a) 50000x and (b) 80000x stacked, few- or multi-layer sheets of **4.1**; (c) SAED confirming few- or multi- layer graphene sheets; (d) selected area from (b) with FFT inset.

The atomic force microscopy (AFM) of **4.1** confirmed the presence of graphene sheets with a wide variety of lateral dimensions from 150 nm to > 5 μm. The thickness of the observed sheets ranged from 0.5 nm to ~2.5 nm (Figure 4.6). The theoretical height of our Phe functional group is ~0.75 nm (not including the aziridino ring linkage), correlating to the addition of 1.5 nm to the thickness of the graphene sheet if functionalization occurs on both sides. Several groups have reported a height of ~1 nm for GO sheets with some variability.[10] Our analysis of **4.1** shows a variety of sheet thicknesses; however, we surmise that the thinnest sheets, 0.5 nm, are only edge functionalized while the thickest at 2.5 nm are basal plane functionalized on both sides. Based on both the TEM and AFM characterization, we conclude that few-layer (n < 5) graphene sheets are produced by this reaction method.

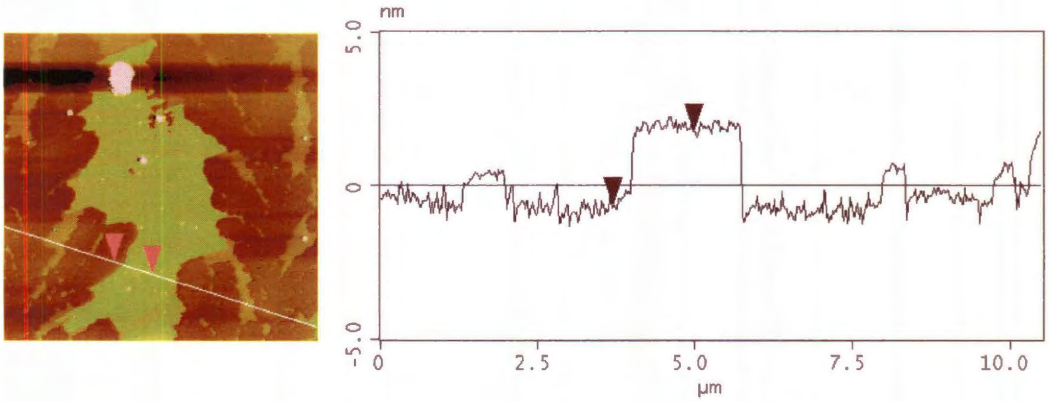


Figure 4.6. The AFM image and associated height profile of Phe-N-μG (**4.1**), spin coated on a cleaved mica substrate from $CHCl_3$ are shown above.

Scanning TEM (STEM) images of **4.1** (Figure 4.7) deposited from DMSO on a lacey carbon grid show few-layer sheets, the majority of which had lengths of over 1 μm and widths of about 500 nm. However, some sheets were as small as 150 x 300 nm, in agreement with TEM and AFM measurements. EDX analysis of the dark spots showed the presence of oxygen in the spots but not nitrogen, therefore, we were unable to definitively confirm that the spots are phenylalaine substituents.

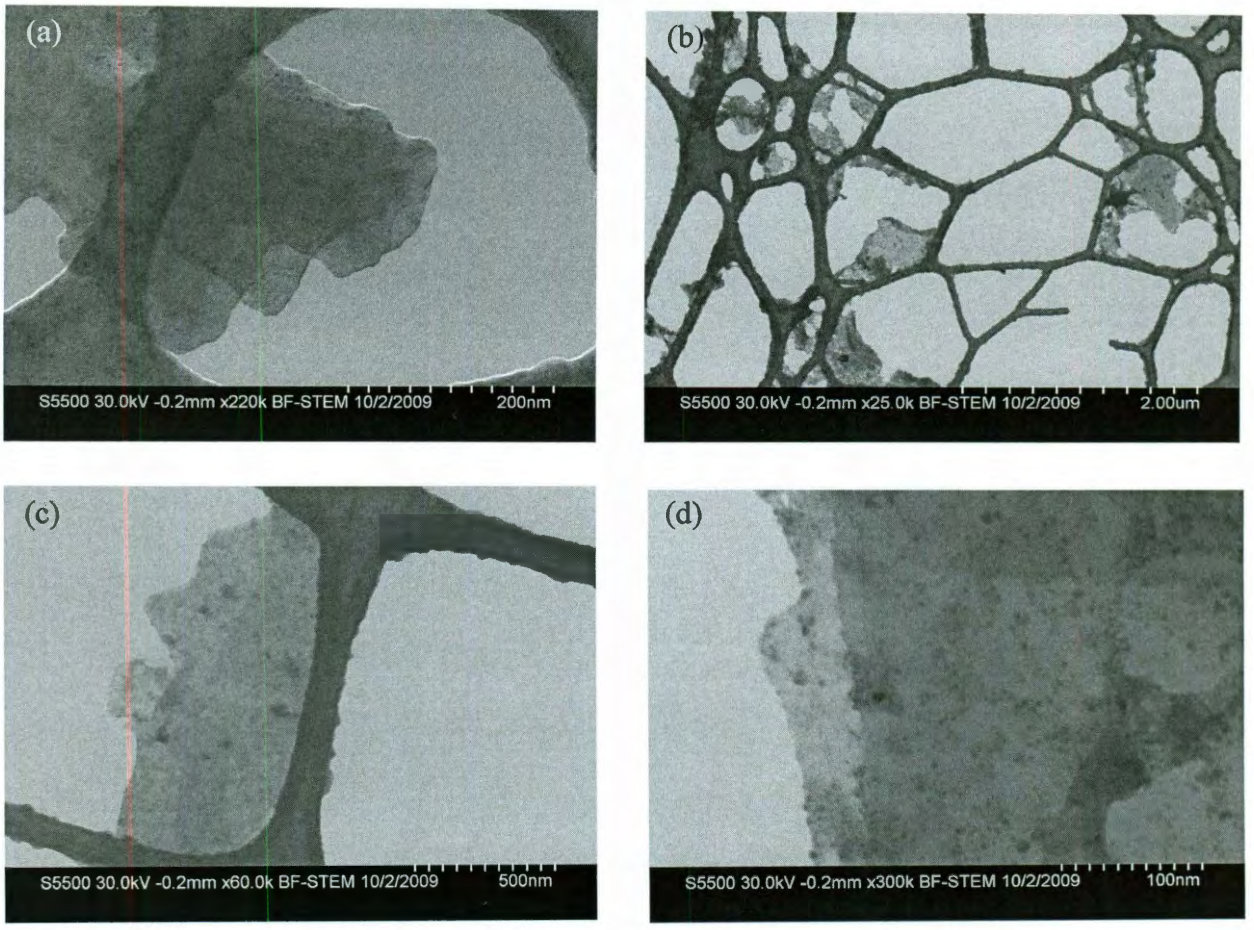


Figure 4.7. Bright Field-STEM images of **4.1** deposited from DMSO showing (a) few layer graphene sheets of width measuring 150, 200, and 275 nm; (b) several large groups of >1 μm single and few layer sheets along with smaller aggregates of flakes; (c) 375 nm x 1.2 μm sheet on top of 1.3 x 0.5 μm sheet; (d) sheet with dark spots covering the basal plane.

The extent of functionalization was obtained from thermogravimetric analysis (TGA) over the temperature range of 130 - 850 °C in agreement with the decomposition of phenylalanine (Figure 4.8). Assuming no mass loss due to μG and taking into account the carbonaceous material remaining from phenylalanine, the final product **4.1** was determined to be 69% phenylalanine (wt.%). This corresponds to one Boc-Phe-N substituent per 10 μG carbons (XPS data shows 13:1 C:Phe-N). It should be noted, however, that each Phe substituent is bonded to two carbon atoms (Scheme 4.1). It has been shown previously that graphene is

preferentially functionalized along it edges.[11] However, by our synthesis method, even with the smallest sheets of **4.1** (150 x 300 nm by STEM), if the functionalization is limited to the edges, only 2-3% by mass of the sample is attributable to Phe-N. Assuming the entire mass loss by TGA is due to covalently bonded Phe, to achieve 69% by mass functionalization of **4.1**, the nitrene addition could not occur solely at the edges of the sheets. STEM images (Figure 4.7) show dark spots along the basal plane of the sheets but EDX analysis of the spots could not definitively confirm that they are Phe. We conclude, therefore, that by our route, functionalization is not limited to the edges and the resulting structure is analogous to that of GO.[4]

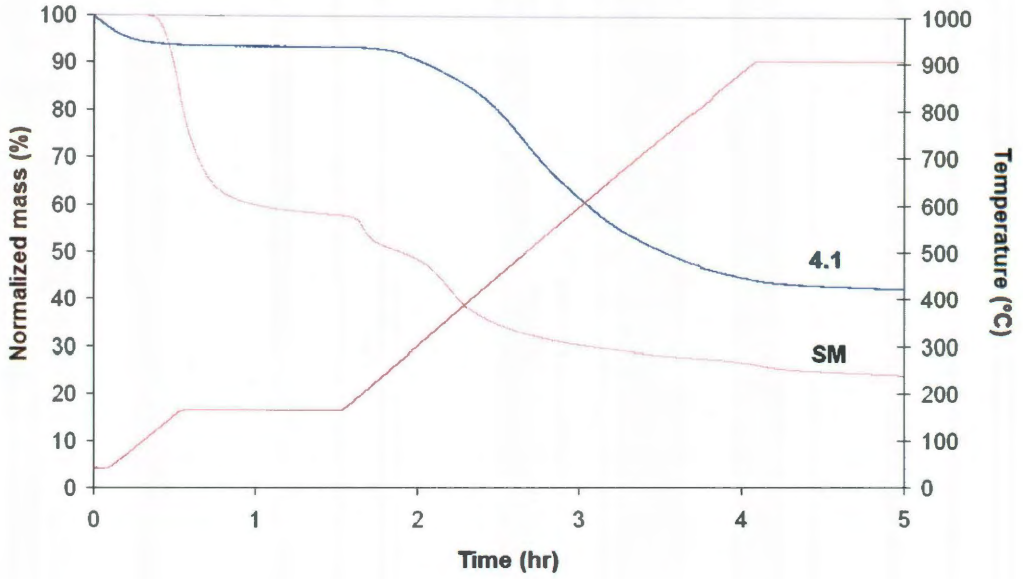


Figure 4.8. The TGA of **4.1** showing 69% normalized mass loss corresponding to a 10:1 C to Boc-Phe-N ratio. Note: 27% of Boc-Phe(N_3)-OH (SM), remains after heating to 850 °C under Ar with a one hour hold time.

Raman spectroscopy (514 nm excitation) of **4.1** showed a D:G ratio of 0.78 in comparison to that of exfoliated μG at 0.66 (Figure 4.9a and b). While there is not a significant change, this could be due to the distribution of functional groups on the surface and/or an effect

of the size of the sheets.[12] The most noticeable feature in the spectrum of **4.1** is the loss of the 2D peak (*ca.* 2700 cm^{-1}). Most previous reports of functionalized graphene omit mention of this peak.[11,13] The only material reported to show a similar loss of the 2D peak is graphene oxide (GO),[14] in which the graphitic structure is almost completely oxidized. Based on the similar spectra of **4.1** to GO (Figure 4.9b and c), the absence of the 2D peak indicates the lack of sp^2 carbons present and, therefore, the high degree of functionalization. Thermolysis of **4.1** at 600 °C for 12 hrs produced a black powder with a D:G ratio of 0.67 and a partially restored 2D peak (Figure 4.9 d) similar to that of the exfoliated μG sample. It should be noted that the bulk μG material prior to exfoliation had a D:G ratio of 0.25 indicating possible modification to the electronic framework simply through sonication in ODCB.[15]

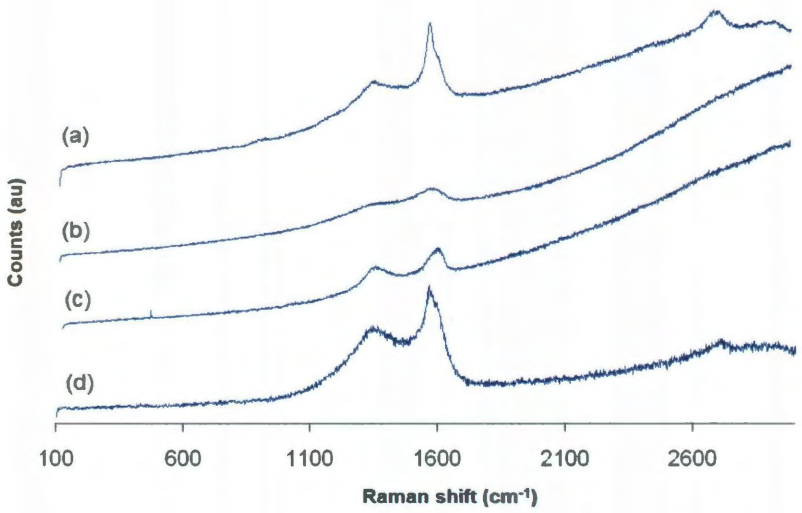


Figure 4.9. The Raman spectra of (a) exfoliated μG, D:G 0.66; (b) 1, D:G 0.78; (c) graphene oxide, D:G 0.77 (d) annealed 1, D:G 0.67 are presented above.

As mentioned previously the XPS spectra of **4.1** did not show the expected 2:1 oxygen to nitrogen ratio that we expected. Given that the Boc protecting group has been shown to be labile at temperatures near the reaction temperature, 180 °C, the concern was that the observed high yield functionalization may be due to polymerization of phenylalanine on the surface of the

graphene. To determine if polymerization of Phe was occurring, we repeated the reaction with a methyl ester of Boc-4-azido-Phe (**4.2**). Protection at both carboxy and amino terminus should eliminate the possibility of peptide formation. When the graphene functionalization was repeated with **4.2**, it was performed at a slightly lower temperature to prevent the loss of the Boc group. The product (**4.3**) was worked up as previously stated. The high resolution XPS spectra of **4.3** showed 13.21% N and 20.58% O, much closer to the 2:1 ratio we expected, but probably still with some loss of the Boc group (Figure 4.10). The nitrogen content as determined by XPS was used to calculate the carbon to Phe-N ratio and confirm the results we obtained by TGA. The carbon to Phe-N ratio of **4.1** and **4.3** were calculated to be 13:1 and 12:1 respectively from XPS data. Given that the degree of functionalization for both samples was similar, even with **4.3** protected at both ends, we felt comfortable with our determination that high yield functionalization of the graphene surface was occurring.

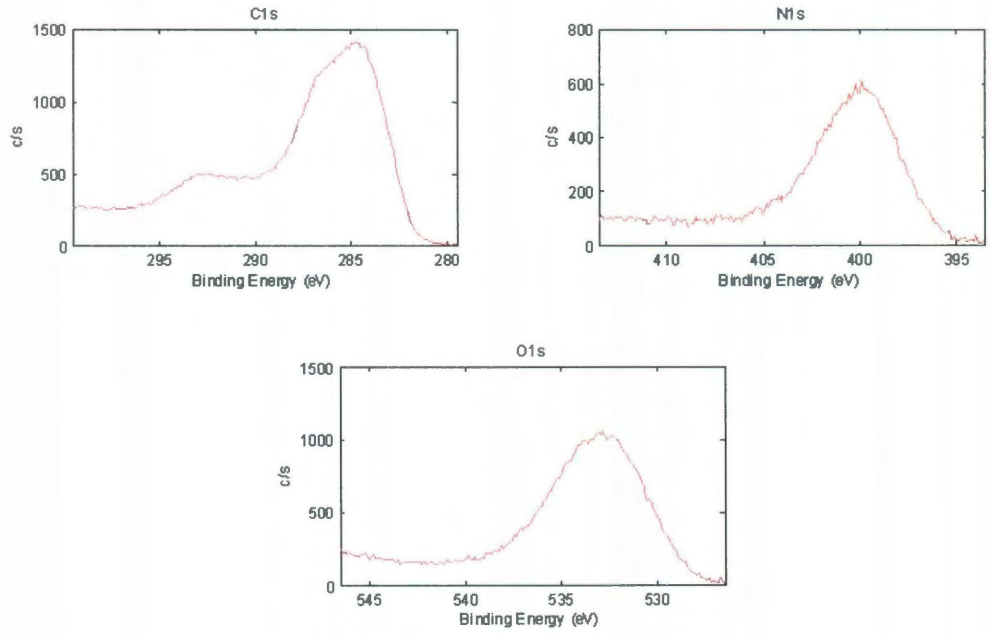


Figure 4.10. The high-resolution XPS spectra of **4.3**, C1s, N1s and O1s spectra (y-axis arb units) showing the presence of C (66.20%), O (20.58%) and N (13.21%). The data corresponds to 12:1 carbon to phenylalanine ratio for the product **4.3**.

Conclusions

In conclusion, we have reported a high yield method of covalently functionalizing exfoliated graphene by nitrene addition in ODCB. The addition results in few-layer, DMSO soluble graphene sheets. Given the extensive range of organic azides accessible by a simple diazo-transfer reaction,[6] this method offers a route to a wide variety of functional addends. The advantages of our method are as follows: 1) no final reduction step is required, 2) a wide variety of functional addends are accessible, 3) reactions are not limited to aqueous/polar conditions, and 4) highly functionalized, few-layer graphene sheets are produced. The control over the extent of functionalization is currently under investigation, as well as the feasibility of **4.1** as a solid phase support for peptide synthesis on the surface of a graphene scaffold.

Experimental

ODCB was purchased from Fisher Scientific and used as received. Microcrystalline synthetic graphite (<20 μm) and Boc-Phe(4-NH_2)-OH and all other reagents were purchased from Sigma-Aldrich and used as received. Sonication was performed with an ultrasonic processor (Cole-Parmer) with probe attachment at 30% amplitude. Centrifugation was at 3300 rpm for 15 min. (Sorvall ST 16, RCF_{3300} = 2045) unless otherwise specified. All reactions are performed under an inert atmosphere.

All XPS data was acquired on a Physical Electronics, Inc. Phi Quantera instrument with monochromated Al-K_α. TEM images were collected with a Jeol JEM-2100F advanced field emission electron microscope with an accelerating voltage of 120 - 200 kV. AFM measurements were obtained using a Digital Instruments NanoScope IIIa scanning probe microscope in tapping mode. AFM samples were spin coated from $CHCl_3$ onto cleaved mica substrates. STEM micrographs were collected with a Hitachi S5500 scanning transmission electron microscope.

ODCB suspensions. Synthetic microcrystalline graphite (50 mg) was added to 20 mL ODCB. The mixture was probe sonicated for one hour. The slurry was then centrifuged for 30

minutes at 4400 rpm. High shear mixing of the ODCB/μG suspension was not performed in an effort to maintain the large lateral dimensions of the μG starting material. Therefore, our dispersions may not be exfoliated to the same degree as previous examples using this method.[7] The supernatant of the μG solution was used in the nitrene reaction. Resuspension and filtration of the μG pellet yielded 41 mg of recovered μG.

Boc-Phe(4-N-μG)-OH (4.1). Boc-Phe(4-N_3)-OH (150 mg in ODCB, 2 - 3 mL) was added to the exfoliated μG solution (~5 mL) in a 50 mL 3-neck flask fitted with a condenser, a thermometer, a stopper and a stir bar. The reaction was diluted to 25 mL with ODCB and the temperature was maintained at 165 °C for 4 days under Ar after which the brown suspension was filtered over a 0.2 um PTFE filter paper. The filter cake was washed copiously with MeOH, $CHCl_3$, and DMF to remove any unreacted Boc-Phe(N_3)-OMe. The fine brown powder was dried in an oven overnight at 125 °C. Final yield: 110 mg.

Boc-Phe(4-N_3)-OMe (4.2). In a 250 mL Shlenk flask, Boc-Phe(4-N_3)-OH (153.6 mg, 0.5 mmol) was dissolved in MeOH (200 mL, seived) under an inert atmosphere. Two equivalents of chlorotrimethylsilane (TMS-Cl, 126.5 μL, 1.0 mmol) were added to the stirring solution. After stirring at room temperature for 5 hours, the reaction mixture was dried on a rotary evaporator. The residue was dissolved in DCM and filtered to remove white precipitate. The filtrate was applied to an SiO_2 plug and eluted with ethyl acetate:hexanes (1:1; R_f 0.85) followed by rotary evaporation. Yield = 43 mg, 0.14 mmol, 27%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.11 [2H, d, *J*(H–H) = 8.35 Hz, C*H*], 6.96 [2H, d, *J*(H–H) = 8.47 Hz, C*H*], 4.98 [1H, d, *J*(H–H) = 7.76 Hz, N*H*], 4.57 [1H, dd, *J*(H–H) = 5.91 Hz, *J*(H–H) = 14.07 Hz, C*H*], 3.71 (3H, s, CH_3), 3.10 [1H, dd, *J*(H–H) = 5.78 Hz, *J*(H–H) = 13.75 Hz, CH_2], 3.01 [1H, dd, *J*(H–H) = 5.94 Hz, *J*(H–H) = 13.75 Hz, CH_2], 1.42 (9H, s, ^{t}Bu). IR (cm^{-1}): ν_{N3} 2115.

Boc-Phe(4-N-μG)-OMe (4.3). Boc-Phe(4-N_3)-OMe (35 mg in ODCB, 2 - 3 mL) was added to the exfoliated μG solution (~5 mL) in a 50 mL 3-neck flask fitted with a condenser, a thermometer, a stopper and a stir bar. The reaction was diluted to 25 mL with ODCB and the temperature was maintained at 165 °C for 4 days under Ar after which the brown suspension was filtered over a 0.2 μm PTFE filter paper. The filter cake was washed copiously with MeOH, $CHCl_3$, and DMF to remove any unreacted Boc-Phe(N_3)-OMe. The fine brown powder was dried in an oven overnight at 125 °C. Final yield: 14 mg.

References

1 (a) K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 10451. (b) A. K. Geim and K. S. Novoselov, *Nature. Mater.*, 2007, 6, 183.

2 K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, *Science*, 2004, **306**, 666.

3 (a) A. B. Bourlinos, D. Gournis, D. Petridis, T. Szabó, A. Szeri, and I. Dékány, *Langmuir*, 2003, **19**, 6050. (b) S. Park and R. S. Ruoff, *Nature Nanotechnology*, 2009, **4**, 217.

4 A. Lerf, H. He, M. Forster, and J. Klinowski, *J. Phys. Chem. B*, 1998, **102**, 4477.

5 (a) M. Prato, C. Li, and F. Wudl, *J. Am. Chem. Soc.*, 1993, **115**, 1148. (b) C. Gao, H. He, L. Zhou, X. Zheng, and Y. Zhang, *Chem. Mater.*, 2009, **21**, 360.

6 E. D. Goddard-Borger and R. V. Stick, *Org. Lett.* 2007, **9**,3797.

7 (a) C. E. Hamilton, J. R. Lomeda, A. Sun, J. M. Tour, and A. R. Barron, *Nano Lett.*, 2009, **9**, 3460. (b) C. E. Hamilton, J. R. Lomeda, Z. Sun, J. M. Tour, and A. R. Barron, *Nano Res.*, 2009, **42**, 473.

8 L. M. Viculis, J. J. Mack, O. M. Mayer, H. T. Hahn, and R. B. Kaner, *J. Mater. Chem.*, 2005, **15**, 974.

[9] V. H. Rawal and M. P. Cava, *Tet. Lett.*, 1985, **26**, 6141.

[10] (a) C. Gomez-Navarro, R. T. Weitz, A. M. Bittner, M. Scolari, A. Mews, M. Burghard, and K. Kern, *Nano. Lett.*, 2007, **7**, 3499. (b) S. Stankovich, R. D. Piner, X. Chen, N. Wu, S. T. Nguyen, and R. S. Ruoff, *J. Mater. Chem.*, 2006, **16**, 155. (c) H. A. Becerril, J. Mao, Z. Liu, R. M. Stoltenberg, Z. Bao, and Y. Chen, *ACS Nano*, 2008, **2**, 463.

[11] J. Chattopadhyay, A. Mukherjee, C. E. Hamilton, J. H. Kang, S. Chakraborty, W. Guo, K. F. Kelly, A. R. Barron, and W. E. Billups, *J. Am. Chem. Soc.*, 2008, **130**, 5414.

[12] F. Tuinstra, and J. L. Koenig, *J. Chem. Phys.*, 1970, **53**, 1126; A. C. Ferrari, *Solid State Commun.*, 2007, **143**, 47.

[13] (a) Y. Liua, J. Zhoua, X. Zhangb, Z. Liub, X. Wana, J. Tianb, T. Wanga, Y. Chen, *Carbon*, 2009, **47**, 3113. (b) S. Ryu, M. Y. Han, J. Maultzsch, T. F. Heinz, P. Kim, M. L. Steigerwald and L. E. Brus, *Nano Lett.*, 2008, **8**, 4597. (c) K. A. Worsley, P. Ramesh, S. K. Mandal, S. Niyogi, M. E. Itkis, and R. C. Haddon, *Chem. Phys. Lett.*, 2007, **445**, 51.

[14] J. R. Lomeda, C. D. Doyle, D. V. Kosynkin, W-F Hwang, and J. M. Tour, *J. Am. Chem. Soc.*, 2008, **130**, 16201.

[15] (a) S. Niyogi, M. A. Hamon, D. E. Perea, C. B. Kang, B. Zhao, S. K. Pal, A. E. Wyant, M. E. Itkis, and R. C. Haddon, *J. Phys. Chem. B.*, 2003, **107**, 8799. (b) K. R. Moonoosawmy and P. Kruse, *J. Phys, Chem. C.*, 2009, **113**, 5133.

Appendix A

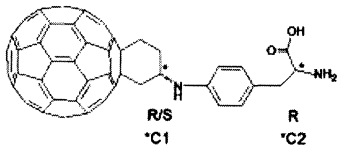
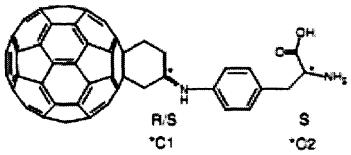
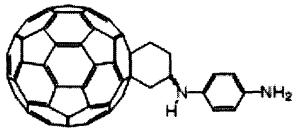
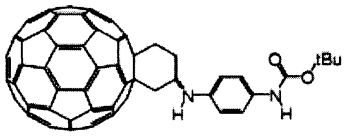
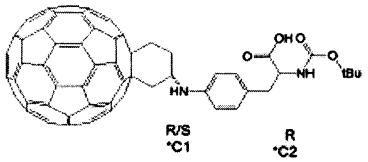
Collaborators Methods and Results

Docking Simulations. The molecular docking simulation was performed using a the crystal structure of HIV-1 PR with the inhibitor haloperidol (PDB code:1AID). All waters of crystallization as determined by the crystal structure remained throughout the docking simulation but were freely rotatable and replaceable by ligand, i.e. inhibitor, docking. Two *in silico* docking algorithms FlexX and Glide Induced Fit Docking (IFD) were used for assessing the binding properties of each compound. A detailed description of the docking simulation is provided elsewhere.[1]

Inhibition Assay. The Ki values were determined by a spectrophotometric assay using the chromogenic peptide substrate LysAlaArgValNle*NphGluAlaNle-NH2 as described by Weber et al.[2] Typically, 8-10 pmol of HIV1-PR was added to 1 mL of 0.1 M sodium acetate buffer (pH 4.7), 0.3 M NaCl, and 4 mM EDTA, containing the substrate at a concentration near the K_m of the enzyme and various concentrations of the inhibitor dissolved in DMSO (from 0.1 mM to 0.01 nM). The final concentrations of DMSO were kept below 2.5% (v/v). Substrate hydrolysis was followed as a decrease in absorbance at 305 nm using a Cary 3 500 UV-vis spectrophotometer (Cambridge, UK). The HIV-1 PR remained stable over the whole reaction time. Inhibition data were analyzed using the equation for competitive inhibition according to Williams and Morrison.[3] The solubility of the compounds was sufficient for all concentrations used in biological and solubility experiments. More specifically, in *in vitro* tests, the DMSO/water mixture has been used as a 1:9 ratio, and no precipitation problems have been observed.

The results of the MD simulations and inhibitory assay are given in Table A.1.

Table A.1. Collaborators results of C_{60} based HIV-1 PR inhibitors.

Structure	Experimental binding affinity, K_i (nM)	Experimental binding energies, (kJ/mol)	Simulated binding energies, Glide/IFD (kJ/mol)
	120 [1]	-39.75	-38.83
	124 [1]	-39.67	-38.28
	6600 [4]	-29.75	-36.65
	> 10000 [4]	> -28.72	-38.45
	437 [4]	-36.52	-39.04
	>10000 [4]	> -28.72	-33.93

References

[1] (a) S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, and M. G. Papadopoulos, *J. Chem. Inf. Model.*, 2009, **49**, 1139. (b) S. Durdagi, T. Mavromoustakos, N. Chronakis, M. G. Papadopoulos, *Bioorg. Med. Chem.*, 2008, **16**, 9957.

[2] J. Weber, J. R. Mesters, M. Lepsik, J. Prejdova, M. Svec, J. Sponarova, P. Mlcochova, K. Skalicka, K. Strisovsky, T. Uhlikova, M. Soucek, L. Machala, M. Stankova, J. Vondrasek, T. Klimkait, H. G. Kraeusslich, R. Hilgenfeld, J. Konvalinka, *J. Mol. Biol.*, 2002, **324**, 739.

[3] J. W. Williams and J. F. Morrison, *Methods Enzymol.*, 1979, **63**, 437.

[4] N. Doostdar, *Doctoral Thesis*, 2010, Rice University.

Appendix B

Publications

1. *In silico* drug screening approach for the design of magic bullets: a successful example with anti-HIV fullerene derivatized amino acids. S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. A. Kumar, A. R Barron, T. Mavromoustakos, and M. G. Papadopoulos, *J. Chem. Info. Model.*, 2009, 49, 1139.

2. Nanoscale enzyme inhibitors: Fullerenes inhibit carbonic anhydrase by occluding the active site entrance . A. Innocenti, S. Durdagi, N. Doostdar, T. A. Strom, A. R. Barron, and C. T. Supuran, *Bioorganic & Medicinal Chemistry*, 2010, 2822.

3. Nitrene addition to exfoliated graphene: a one-step route to highly functionalized graphene. T. A. Strom, E. P. Dillon, C. E. Hamilton and A. R. Barron, *Chem. Commun.*, 2010, 46, 4097.

4. A simple quick route to fullerene amino acid derivatives. T. A. Strom and A. R. Barron, *Chem. Commun.*, 2010, 46, 4764.